

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:21:09 ; Search time 48.7059 Seconds  
(without alignments)  
400.276 Million cell updates/sec

Title: US-09-620-955B-9  
Perfect score: 379  
Sequence: 1 LVPRGSVSTHHHHHQQQQQ.....HHGNSGPPEFPGRLERPHRD 69

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	379	100.0	69	4	AAB69604	Aab69604 Huntingti
2	245	64.6	113	4	AAB69615	Aab69615 Huntingti
3	212	55.9	86	2	AAW95073	Aaw95073 GST-HD fu
4	212	55.9	86	2	AAW95078	Aaw95078 GST-HD fu
5	199.5	52.6	3502	4	ABB58382	Abb58382 Drosophil
6	197	52.0	325	4	ABB71488	Abb71488 Drosophil
7	192	50.7	171	5	AAE26650	Aae26650 Human hun
8	191	50.4	94	2	AAW95075	Aaw95075 GST-HD fu
9	191	50.4	94	2	AAW95080	Aaw95080 GST-HD fu

10	189.5	50.0	821	4	AAM78983	Aam78983	Human pro
11	189	49.9	905	5	ABG93053	Abg93053	S. cerevi
12	189	49.9	905	6	ABR53130	Abr53130	Protein s
13	188.5	49.7	910	2	AAAY22191	Aay22191	Mouse bra
14	188.5	49.7	910	5	ABJ10802	Abj10802	Mouse HCN
15	187	49.3	108	2	AAW95071	Aaw95071	Amino aci
16	187	49.3	108	2	AAW95076	Aaw95076	Amino aci
17	186	49.1	611	5	ABG93274	Abg93274	C. albica
18	186	49.1	1057	4	ABB62880	Abb62880	Drosophil
19	185.5	48.9	843	4	AAM79967	Aam79967	Human pro
20	185	48.8	149	4	AAB72673	Aab72673	Polygluta
21	185	48.8	692	4	ABB64835	Abb64835	Drosophil
22	183	48.3	2280	4	ABB61650	Abb61650	Drosophil
23	182.5	48.2	722	7	ADC55550	Adc55550	Mutant hu
24	182.5	48.2	951	7	ADC55544	Adc55544	Mutant hu
25	182	48.0	915	6	ABU40496	Abu40496	Protein e
26	182	48.0	1090	5	ABP35656	Abp35656	Fungal ZB
27	182	48.0	1761	4	ABB59512	Abb59512	Drosophil
28	181	47.8	98	4	AAB69610	Aab69610	Huntingti
29	180.5	47.6	758	4	AAB30812	Aab30812	Amino aci
30	180.5	47.6	2023	3	AAAY54320	Aay54320	Amino aci
31	180	47.5	64	4	AAB69607	Aab69607	Huntingti
32	180	47.5	89	4	AAB69608	Aab69608	Huntingti
33	180	47.5	121	4	AAB69609	Aab69609	Huntingti
34	180	47.5	123	4	AAB69611	Aab69611	Huntingti
35	180	47.5	145	4	AAB69614	Aab69614	Huntingti
36	180	47.5	155	4	AAB69612	Aab69612	Huntingti
37	179	47.2	80	4	AAB69622	Aab69622	TATA bind
38	179	47.2	128	5	ABP53102	Abp53102	Human Zac
39	179	47.2	338	5	AAU77921	Aau77921	Human Tat
40	179	47.2	371	2	AAW73369	Aaw73369	Epitope t
41	179	47.2	821	4	ABB70509	Abb70509	Drosophil
42	179	47.2	1186	4	ABB63516	Abb63516	Drosophil
43	179	47.2	2150	5	AAO22566	Aao22566	Wooden le
44	178.5	47.1	666	4	ABB58019	Abb58019	Drosophil
45	178.5	47.1	861	4	ABB62534	Abb62534	Drosophil

#### ALIGNMENTS

##### RESULT 1

AAB69604

ID AAB69604 standard; protein; 69 AA.

XX

AC AAB69604;

XX

DT 30-APR-2001 (first entry)

XX

DE Huntingtin accumulation inhibitor peptide GST-DRPLA-Q35.

XX

KW Neurological disorder; Huntington's disease; Alzheimer's disease;

KW Parkinson's disease; prion disease; frontotemporal dementia;

KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;

KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;

KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.

XX

OS Synthetic.  
 XX  
 PN WO200106989-A2.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 24-JUL-2000; 2000WO-US020131.  
 XX  
 PR 27-JUL-1999; 99US-0146047P.  
 PR 21-JUL-2000; 2000US-00620955.  
 XX  
 PA (HUST/) HUSTON J S.  
 PA (MESS/) MESSER A.  
 PA (LECE/) LECERF J.  
 XX  
 PI Huston JS, Messer A, Lecerf J;  
 XX  
 DR WPI; 2001-182700/18.  
 XX  
 PT Inhibiting intracellular polypeptide accumulation, useful for treating  
 PT neurological disorders, e.g. Alzheimer's disease, comprises contacting  
 PT the polypeptide with a specific intrabody.  
 XX  
 PS Disclosure; Page 96; 108pp; English.  
 XX  
 CC The present invention describes a method for inhibiting the formation of  
 CC aggregates of certain proteins, involving contacting the protein with a  
 CC binding molecule known as an intrabody. Proteins to be bound include  
 CC those associated with neurological disorders, and so the method can be  
 CC used in the prevention of diseases such as Alzheimer's, Parkinson's and  
 CC Huntington's diseases, prion disease, frontotemporal dementia,  
 CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,  
 CC dentatorubal-pallidoluytian atrophy, spinocerebellar ataxia type 1  
 CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7  
 XX  
 SQ Sequence 69 AA;  
  
 Query Match 100.0%; Score 379; DB 4; Length 69;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-35;  
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 Qy 1 LVPRGSVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFP 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 LVPRGSVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFP 60  
  
 Qy 61 GRLERPHRD 69  
 ||||||||  
 Db 61 GRLERPHRD 69  
  
 RESULT 2  
 AAB69615  
 ID AAB69615 standard; protein; 113 AA.  
 XX  
 AC AAB69615;  
 XX  
 DT 30-APR-2001 (first entry)

XX Huntingtonin accumulation inhibitor peptide GFP-DRPLA-Q81.  
XX  
KW Neurological disorder; Huntington's disease; Alzheimer's disease;  
KW Parkinson's disease; prion disease; frontotemporal dementia;  
KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;  
KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;  
KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.  
XX  
OS Synthetic.  
XX  
PN WO200106989-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 24-JUL-2000; 2000WO-US020131.  
XX  
PR 27-JUL-1999; 99US-0146047P.  
PR 21-JUL-2000; 2000US-00620955.  
XX  
PA (HUST/) HUSTON J S.  
PA (MESS/) MESSER A.  
PA (LECE/) LECERF J.  
XX  
PI Huston JS, Messer A, Lecerf J;  
XX  
DR WPI; 2001-182700/18.  
XX  
PT Inhibiting intracellular polypeptide accumulation, useful for treating  
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting  
PT the polypeptide with a specific intrabody.  
XX  
PS Disclosure; Page 100; 108pp; English.  
XX  
CC The present invention describes a method for inhibiting the formation of  
CC aggregates of certain proteins, involving contacting the protein with a  
CC binding molecule known as an intrabody. Proteins to be bound include  
CC those associated with neurological disorders, and so the method can be  
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and  
CC Huntington's diseases, prion disease, frontotemporal dementia,  
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,  
CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1  
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7  
XX  
SQ Sequence 113 AA;

```

Qy      5  GSVSTHHHHH-----QQQQ 18
          |||||
Db      15  GSVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 74

Qy      19  QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEF 59
          |||||
Db      75  QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH--SGPPEF 113

```

RESULT 3

AAW95073

ID AAW95073 standard; protein; 86 AA.

XX

AC AAW95073;

XX

DT 20-MAY-1999 (first entry)

XX

DE GST-HD fusion protein GST-HD51DELP.

XX

KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;  
KW polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;  
KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;  
KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;  
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.

XX

OS Synthetic.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 1

FT /note= "this residue is connected to a GST protein which  
FT is not indicated in the sequence"

XX

PN WO9906838-A2.

XX

PD 11-FEB-1999.

XX

PF 31-JUL-1998; 98WO-EP004810.

XX

PR 01-AUG-1997; 97EP-00113320.

XX

PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX

PI Wanker E, Lehrach H, Scherzinger E, Bates G;

XX

DR WPI; 1999-153955/13.

XX

PT Detecting amyloid-like fibrils or protein aggregates insoluble in  
PT detergent or urea - from their retention on a filter, used for diagnosis,  
PT particularly of diseases associated with polyglutamine expansion.

XX

PS Disclosure; Fig 8; 56pp; English.

XX

CC The invention relates to the detection of amyloid-like fibrils or protein  
CC aggregates, insoluble in detergents or urea. The method comprises: (a)  
CC applying material suspected of containing protein aggregates to a filter;  
CC and (b) detecting retention of protein aggregates on the filter. This  
CC method also helps to identify inhibitors of protein aggregates formation.  
CC The method is particularly used to detect protein aggregates that are  
CC indicative of disease, for assessing onset or progression of the  
CC diseases. The inhibitors identified are potential therapeutic agents for  
CC treating the diseases. Other applications include detection of inclusion  
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases  
CC associated with polyglutamine expansion are particularly diagnosed, e.g.

CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar  
 CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II  
 CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;  
 CC scrapie. The protein aggregates can now be detected simply, routinely and  
 CC rapidly, without requiring sophisticated equipment. The method can be  
 CC made quantitative, by analysing a series of dilutions, and can be  
 CC automated to allow many samples to be analysed on the same filter.  
 CC Sequences AAW95072-75 represent GST-HD fusion proteins  
 XX  
 SQ Sequence 86 AA;

Query Match 55.9%; Score 212; DB 2; Length 86;  
 Best Local Similarity 68.8%; Pred. No. 1.2e-16;  
 Matches 44; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

Qy 6 SVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPGRLER 65  
 | | | | | | | | | | | | | | | | | | | | | | : | | |  
 Db 23 SFQQPPPLER 82  
 Qy 66 PHRD 69  
 | | | |  
 Db 83 PHRD 86

RESULT 4  
 AAW95078

ID AAW95078 standard; protein; 86 AA.  
 XX  
 AC AAW95078;  
 XX  
 DT 20-MAY-1999 (first entry)  
 XX  
 DE GST-HD fusion protein GST-HD51DELP.  
 XX  
 KW Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;  
 KW protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;  
 KW Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;  
 KW dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;  
 KW GST-HD; HD.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1  
 FT /note= "this residue is connected to a GST protein which  
 FT is not indicated in the sequence"  
 XX  
 PN WO9906545-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-EP004811.  
 XX  
 PR 01-AUG-1997; 97EP-00113306.  
 XX  
 PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.



OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL02485.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 1938; 2lpp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 3502 AA;

Query Match 52.6%; Score 199.5; DB 4; Length 3502;  
 Best Local Similarity 78.4%; Pred. No. 1.1e-13;  
 Matches 40; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

Qy 11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPG 61  
 |||||||||||||||||||||:|||| |||| | || ||  
 Db 220 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQKQQQHHMQQQQQQQPLS-PPHPPG 269

# RESULT 6

ABB71488

ID ABB71488 standard; protein; 325 AA.

XX

AC ABB71488;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 41256.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;





XX  
 KW Human; protein misfolding; Alzheimer's disease; AD; Parkinson's disease;  
 KW PD; Familial amyloid polyneuropathy; tauopathy; frontotemporal dementia;  
 KW Pick disease; lobar atrophy; trinucleotide disease; fragile-X syndrome;  
 KW Huntington's disease; spinocerebellar ataxia; SCA; myotonic dystrophy;  
 KW dentatorubral pallidoluysian atrophy; DRPLA; Creutzfeldt-Jacob disease;  
 KW CJD; prion disease; Gerstmann-Straussler-Scheinker disease; GSS; FFI;  
 KW fatal familia insomnia; mad cow disease; scrapie; kuru; anticonvulsant;  
 KW nootropic; neuroprotective; cerebroprotective; htQ103 protein.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200265136-A2.  
 XX  
 PD 22-AUG-2002.  
 XX  
 PF 15-FEB-2002; 2002WO-US004632.  
 XX  
 PR 15-FEB-2001; 2001US-0269157P.  
 XX  
 PA (UYCH-) UNIV CHICAGO.  
 XX  
 PI Lindquist S, Krobitsch S, Outeiro T;  
 XX  
 DR WPI; 2002-667026/71.  
 DR N-PSDB; AAD44410.  
 XX  
 PT Screening for therapeutic agents for protein misfolding disease, by  
 PT contacting a yeast cell with compound, that expresses misfolded disease  
 PT protein, and with a toxicity inducing agent, and evaluating cell for  
 PT viability.  
 XX  
 PS Disclosure; Page 88; 93pp; English.  
 XX  
 CC The present invention relates to novel screening methods for identifying  
 CC therapeutic agents for diseases associated with protein misfolding. The  
 CC method involves contacting a yeast cell with a candidate compound, where  
 CC the yeast cell expresses a polypeptide comprising a misfolded disease  
 CC protein, contacting the yeast cell with a toxicity inducing agent and  
 CC evaluating the yeast cell for viability, where the viability indicates  
 CC the candidate compound is a candidate therapeutic agent. The method is  
 CC useful to screen for therapeutic agents for diseases associated with  
 CC protein misfolding such as Alzheimer's disease (AD), Parkinson's disease  
 CC (PD), Familial amyloid polyneuropathy, tauopathies (e.g. Pick disease,  
 CC lobar atrophy, frontotemporal dementia) or trinucleotide diseases (e.g.  
 CC Huntington's disease, spinocerebellar ataxia (SCA), fragile-X syndrome,  
 CC myotonic dystrophy, dentatorubral pallidoluysian atrophy (DRPLA) and  
 CC prion diseases (e.g. Creutzfeldt-Jacob disease (CJD), fatal familia  
 CC insomnia (FFI), Gerstmann-Straussler-Scheinker disease (GSS), mad cow  
 CC disease, scrapie and kuru). The method is useful for treating a patient  
 CC with Huntington's disease or Parkinson's disease. The present sequence is  
 CC human huntington (htQ103) protein. This sequence is used to illustrate  
 CC the method of the invention  
 XX  
 SQ Sequence 171 AA;

Query Match

50.7%; Score 192; DB 5; Length 171;

Best Local Similarity 76.9%; Pred. No. 4.4e-14;  
Matches 40; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 15 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPGRLERP 66  
||||| :| :|  
Db 85 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQP 136

RESULT 8

AAW95075

ID AAW95075 standard; protein; 94 AA.

XX

AC AAW95075;

XX

DT 20-MAY-1999 (first entry)

XX

DE GST-HD fusion protein GST-HD51DELPBio.

XX

KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;  
KW polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;  
KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;  
KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;  
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.

XX

OS Synthetic.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 1

FT /note= "this residue is connected to a GST protein which  
FT is not indicated in the sequence"

XX

PN WO9906838-A2.

XX

PD 11-FEB-1999.

XX

PF 31-JUL-1998; 98WO-EP004810.

XX

PR 01-AUG-1997; 97EP-00113320.

XX

PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX

PI Wanker E, Lehrach H, Scherzinger E, Bates G;

XX

DR WPI; 1999-153955/13.

XX

PT Detecting amyloid-like fibrils or protein aggregates insoluble in  
PT detergent or urea - from their retention on a filter, used for diagnosis,  
PT particularly of diseases associated with polyglutamine expansion.

XX

PS Disclosure; Fig 8; 56pp; English.

XX

CC The invention relates to the detection of amyloid-like fibrils or protein  
CC aggregates, insoluble in detergents or urea. The method comprises: (a)  
CC applying material suspected of containing protein aggregates to a filter;  
CC and (b) detecting retention of protein aggregates on the filter. This  
CC method also helps to identify inhibitors of protein aggregates formation.

CC The method is particularly used to detect protein aggregates that are  
 CC indicative of disease, for assessing onset or progression of the  
 CC diseases. The inhibitors identified are potential therapeutic agents for  
 CC treating the diseases. Other applications include detection of inclusion  
 CC bodies in bacteria and to study kinetics of aggregate formation. Diseases  
 CC associated with polyglutamine expansion are particularly diagnosed, e.g.  
 CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar  
 CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II  
 CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;  
 CC scrapie. The protein aggregates can now be detected simply, routinely and  
 CC rapidly, without requiring sophisticated equipment. The method can be  
 CC made quantitative, by analysing a series of dilutions, and can be  
 CC automated to allow many samples to be analysed on the same filter.  
 CC Sequences AAW95072-75 represent GST-HD fusion proteins

XX

SQ Sequence 94 AA;

Query Match 50.4%; Score 191; DB 2; Length 94;  
 Best Local Similarity 78.0%; Pred. No. 3.2e-14;  
 Matches 39; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 15 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPGRLE 64  
 |||||  
 Db 36 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQPPPLEGIFE 85

# RESULT 9

AAW95080

ID AAW95080 standard; protein; 94 AA.

XX

AC AAW95080;

XX

DT 20-MAY-1999 (first entry)

XX

DE GST-HD fusion protein GST-HD51DELPBio.

XX

KW Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;  
 KW protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;  
 KW Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;  
 KW dentatorubral pallidoluyian atrophy; Creutzfeld-Jakob disease; enzyme;  
 KW GST-HD; HD.

XX

OS Synthetic.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 1

FT /note= "this residue is connected to a GST protein which  
 FT is not indicated in the sequence"

XX

PN WO9906545-A2.

XX

PD 11-FEB-1999.

XX

PF 31-JUL-1998; 98WO-EP004811.

XX

PR 01-AUG-1997; 97EP-00113306.



OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US004098.  
 XX  
 PR 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR N-PSDB; AAK52116.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX  
 PS Claim 20; Page 3982-3984; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX  
 SQ Sequence 821 AA;

Query Match 50.0%; Score 189.5; DB 4; Length 821;  
 Best Local Similarity 71.9%; Pred. No. 3.8e-13;  
 Matches 41; Conservative 3; Mismatches 12; Indels 1; Gaps 1;

Qy 1 LVPRGSVS-THHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGP 56  
 | |::| | | :| | | | | | | | | | | | | | | | | | | | | | | : |  
 Db 181 LANMGSLSQTPGHKAEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHL SRAP 237

RESULT 11

ABG93053

ID ABG93053 standard; protein; 905 AA.

XX

AC ABG93053;

XX

DT 21-NOV-2002 (first entry)

XX

DE *S. cerevisiae* BAX-associated protein fragment SEQ ID 64.

XX

KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;  
KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;  
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;  
KW neurodegeneration; cell death.

XX

OS *Saccharomyces cerevisiae*.

XX

PN WO200264766-A2.

XX

PD 22-AUG-2002.

XX

PF 21-DEC-2001; 2001WO-EP015398.

XX

PR 22-DEC-2000; 2000EP-00870318.

PR 04-JAN-2001; 2001EP-00870002.

PR 09-JAN-2001; 2001EP-00870003.

XX

PA (JANC ) JANSSEN PHARM NV.

XX

PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;

XX

DR WPI; 2002-667002/71.

DR N-PSDB; ABQ76319.

XX

PT New isolated nucleic acid representing a synthetic BAX-gene, useful as  
PT medicament for treating, preventing and/or alleviating yeast or fungal  
PT infections or proliferative disorders, or for preventing apoptosis in  
PT certain diseases.

XX

PS Claim 36; Fig 1; 344pp; English.

XX

CC This invention describes a novel nucleic acid representing a synthetic  
CC Bax gene. The Bax gene of the invention is useful for identifying Bax-  
CC resistant yeast or fungi, identifying, or obtaining and identifying  
CC *Candida* spp. sequences that are differentially expressed in a pathway  
CC eventually leading to programmed cell death or identifying inhibitors or  
CC inhibitor sequences of Bax-induced cell death. The products of the  
CC invention have cytostatic, fungicide; immunosuppressive, virucide and  
CC vasotropic activity and can be used in vaccines or for gene therapy. The  
CC isolated nucleic acids, polypeptides, pharmaceutical compositions,  
CC antisense molecules and antibodies are useful as medicaments or in  
CC preparing a medicament for treating, preventing and/or alleviating  
CC diseases associated with yeast or fungi or proliferative disorders, such  
CC as cancer, or for preventing apoptosis in certain diseases. The compounds  
CC or polypeptides, or the genetically modified organism are useful for  
CC preparing a medicament for modifying the endogenic flora of humans and  
CC other mammals. The vaccine is useful for immunising against yeast or  
CC fungal infections. Apoptosis-related diseases include autoimmune disease,











KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;  
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 1  
FT /note= "GST protein connected to the N-terminal"  
FT Misc-difference 25  
FT /note= "polyglutamine expansion that can comprise upto 51  
FT glutamines"  
XX  
PN WO9906838-A2.  
XX  
PD 11-FEB-1999.  
XX  
PF 31-JUL-1998; 98WO-EP004810.  
XX  
PR 01-AUG-1997; 97EP-00113320.  
XX  
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX  
PI Wanker E, Lehrach H, Scherzinger E, Bates G;  
XX  
DR WPI; 1999-153955/13.  
XX  
PT Detecting amyloid-like fibrils or protein aggregates insoluble in  
PT detergent or urea - from their retention on a filter, used for diagnosis,  
PT particularly of diseases associated with polyglutamine expansion.  
XX  
PS Example 1; Fig 2; 56pp; English.  
XX  
CC The invention relates to the detection of amyloid-like fibrils or protein  
CC aggregates, insoluble in detergents or urea. The method comprises: (a)  
CC applying material suspected of containing protein aggregates to a filter;  
CC and (b) detecting retention of protein aggregates on the filter. This  
CC method also helps to identify inhibitors of protein aggregates formation.  
CC The method is particularly used to detect protein aggregates that are  
CC indicative of disease, for assessing onset or progression of the  
CC diseases. The inhibitors identified are potential therapeutic agents for  
CC treating the diseases. Other applications include detection of inclusion  
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases  
CC associated with polyglutamine expansion are particularly diagnosed, e.g.  
CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar  
CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II  
CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;  
CC scrapie. The protein aggregates can now be detected simply, routinely and  
CC rapidly, without requiring sophisticated equipment. The method can be  
CC made quantitative, by analysing a series of dilutions, and can be  
CC automated to allow many samples to be analysed on the same filter. The  
CC present sequence represents the Huntington's gene exon 1 translation  
CC product which is connected to a GST protein to form a fusion protein. The  
CC sequence of the GST protein is not indicated  
XX  
SQ Sequence 108 AA;

Query Match 49.3%; Score 187; DB 2; Length 108;  
Best Local Similarity 76.5%; Pred. No. 1e-13;  
Matches 39; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

```
Qy      16 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPGRLERP 66
          |||||||||||||||||||||||||||||||||  ||  | :| :|
Db      25 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQP 75
```

Search completed: March 12, 2004, 15:38:29  
Job time : 50.7059 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:38:34 ; Search time 14.2059 Seconds  
(without alignments)  
250.755 Million cell updates/sec

Title: US-09-620-955B-9  
Perfect score: 379  
Sequence: 1 LVPRGSVSTHHHHHQQQQQ.....HHGNSGPPEFPGRLERPHRD 69

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query		Match Length DB	ID	Description
	Score				
1	188.5	49.7	910 4	US-08-997-685A-2	Sequence 2, Appli
2	180.5	47.6	2023 4	US-09-491-356C-8	Sequence 8, Appli
3	177	46.7	2074 4	US-09-491-356C-9	Sequence 9, Appli
4	171.5	45.3	1420 4	US-09-125-635-4	Sequence 4, Appli
5	167	44.1	816 2	US-08-267-803B-9	Sequence 9, Appli
6	167	44.1	816 3	US-09-041-886-17	Sequence 17, Appl
7	161	42.5	1184 4	US-09-266-225D-18	Sequence 18, Appl
8	161	42.5	1185 3	US-09-041-886-23	Sequence 23, Appl
9	158.5	41.8	729 4	US-09-625-188-20	Sequence 20, Appl
10	155	40.9	542 1	US-07-814-964-13	Sequence 13, Appl
11	155	40.9	542 1	US-08-258-442-13	Sequence 13, Appl

12	155	40.9	542	1	US-08-328-809-8	Sequence 8, Appli
13	155	40.9	542	4	US-08-866-840-8	Sequence 8, Appli
14	155	40.9	542	5	PCT-US92-11107-13	Sequence 13, Appl
15	155	40.9	2703	1	US-08-185-432-19	Sequence 19, Appl
16	155	40.9	2703	4	US-08-899-232-4	Sequence 4, Appli
17	151	39.8	788	2	US-08-918-914-4	Sequence 4, Appli
18	151	39.8	1282	4	US-09-543-681A-5419	Sequence 5419, Ap
19	149	39.3	678	5	PCT-US93-03027-3	Sequence 3, Appli
20	148.5	39.2	591	3	US-08-965-903B-2	Sequence 2, Appli
21	147	38.8	528	4	US-09-086-663A-82	Sequence 82, Appl
22	147	38.8	548	4	US-09-086-663A-71	Sequence 71, Appl
23	147	38.8	596	4	US-09-086-663A-2	Sequence 2, Appli
24	147	38.8	596	4	US-09-086-663A-80	Sequence 80, Appl
25	147	38.8	1402	4	US-09-125-635-12	Sequence 12, Appl
26	144.5	38.1	360	2	US-08-531-927B-2	Sequence 2, Appli
27	144.5	38.1	360	3	US-09-041-886-13	Sequence 13, Appl
28	144	38.0	303	1	US-08-185-432-5	Sequence 5, Appli
29	144	38.0	737	1	US-08-185-432-2	Sequence 2, Appli
30	144	38.0	737	1	US-08-185-432-4	Sequence 4, Appli
31	143	37.7	428	1	US-08-190-802A-29	Sequence 29, Appl
32	143	37.7	428	3	US-08-477-346-29	Sequence 29, Appl
33	143	37.7	428	4	US-08-473-089-29	Sequence 29, Appl
34	143	37.7	428	4	US-08-487-072A-29	Sequence 29, Appl
35	142	37.5	513	3	US-09-100-193-3	Sequence 3, Appli
36	139	36.7	71	4	US-09-146-054-9	Sequence 9, Appli
37	139	36.7	71	4	US-09-664-977A-9	Sequence 9, Appli
38	138.5	36.5	538	4	US-09-457-040B-23	Sequence 23, Appl
39	136	35.9	205	4	US-09-134-000C-4540	Sequence 4540, Ap
40	135	35.6	1507	4	US-09-914-259-37	Sequence 37, Appl
41	134.5	35.5	903	2	US-08-853-310-2	Sequence 2, Appli
42	134	35.4	546	4	US-09-457-040B-24	Sequence 24, Appl
43	134	35.4	1003	4	US-09-521-511C-11	Sequence 11, Appl
44	134	35.4	1088	4	US-09-233-857-13	Sequence 13, Appl
45	134	35.4	1099	4	US-09-442-100-2	Sequence 2, Appli

#### ALIGNMENTS

##### RESULT 1

US-08-997-685A-2

; Sequence 2, Application US/08997685A

; Patent No. 6551821

; GENERAL INFORMATION:

; APPLICANT: The Trustees of Columbia University

; APPLICANT: Kandel, Eric

; TITLE OF INVENTION: Brain Cyclic Nucleotide Gated Ion Channel and Uses  
Thereof

; FILE REFERENCE: 0575/54806

; CURRENT APPLICATION NUMBER: US/08/997,685A

; CURRENT FILING DATE: 1997-12-12

; NUMBER OF SEQ ID NOS: 60

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 910

; TYPE: PRT

; ORGANISM: mouse







RESULT 4

US-09-125-635-4

```
; Sequence 4, Application US/09125635
; Patent No. 6562589
; GENERAL INFORMATION:
; APPLICANT: THE UNITED STATES OF AMERICA represented by THE SE
; TITLE OF INVENTION: AIB1, A novel steroid receptor co-activator
; FILE REFERENCE: 49944
; CURRENT APPLICATION NUMBER: US/09/125,635
; CURRENT FILING DATE: 1998-08-21
; PRIOR APPLICATION NUMBER: 60/049,728
; PRIOR FILING DATE: 1997-06-17
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1420
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-125-635-4
```

```
Query Match          45.3%; Score 171.5; DB 4; Length 1420;
Best Local Similarity 73.5%; Pred. No. 8.6e-12;
Matches 36; Conservative 2; Mismatches 8; Indels 3; Gaps 1;
```

```
Qy      12 HHHQQQQ---QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPP 57
      || :||: |||||
Db      1232 HHFRQQRVAMMMQQQQQQQQQQQQQQQQQQQQQQQQQQQQQTQAFSPPP 1280
```

RESULT 5

US-08-267-803B-9

```
; Sequence 9, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muetting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55458-1415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,803B
```

```

; FILING DATE: 28-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McCormack, Myra H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00030120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-267-803B-9

```

## RESULT 6

```

US-09-041-886-17
; Sequence 17, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

```

```
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-17
```

```
Query Match          44.1%; Score 167; DB 3; Length 816;
Best Local Similarity 50.0%; Pred. No. 1.6e-11;
Matches 42; Conservative 4; Mismatches 20; Indels 18; Gaps 2;
```

```
Qy      1 LVPRGSVS-THHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ----- 45
      |  ||:| | | :| | | | | | | | | | | | | | | | | | | | | | |
Db      181 LANMGSLSQTPGHKAEQQQQQQQQQQQQQHQQHQQQQQQQQQQQQQQQHLSRAPGLITPGSP 240

Qy      46 ---QQQQHHGNSGPPEFPGRLERP 66
      || |: | |: || |
Db      241 PPAQQNQYVHISSSPQNTGRTASP 264
```

#### RESULT 7

```
US-09-266-225D-18
; Sequence 18, Application US/09266225D
; Patent No. 6573364
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishan
; APPLICANT: Kingsmore, Stephen
; APPLICANT: Tchernev, Velizar
; TITLE OF INVENTION: Isolation and Characterization of Hermansky-Pudlak
; TITLE OF INVENTION: Syndrome (HPS) Protein Complexes and HPS Protein-
; TITLE OF INVENTION: Interacting Proteins
; FILE REFERENCE: 15966-523
; CURRENT APPLICATION NUMBER: US/09/266,225D
; CURRENT FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-266-225D-18
```

```
Query Match          42.5%; Score 161; DB 4; Length 1184;
Best Local Similarity 55.0%; Pred. No. 1.2e-10;
Matches 33; Conservative 0; Mismatches 5; Indels 22; Gaps 2;
```

```
Qy      7 VSTHHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPGRLERP 66
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      475 VSTHHHHH-----QQQQQQQQQQQQQQHHGNSGPPP-PGAFPHP 512
```

RESULT 8

US-09-041-886-23

```
; Sequence 23, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
;   APPLICANT: Bredesen, Dale E.
;   APPLICANT: Rabizadeh, Sharroz
;   TITLE OF INVENTION: Proapoptotic Peptides, Dependence
;   TITLE OF INVENTION: Polypeptides and Methods of Use
;   NUMBER OF SEQUENCES: 72
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Campbell & Flores LLP
;     STREET: 4370 La Jolla Village Drive, Suite 700
;     CITY: San Diego
;     STATE: California
;     COUNTRY: United States
;     ZIP: 92122
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/041,886
;     FILING DATE:
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Campbell, Cathryn A.
;     REGISTRATION NUMBER: 31,815
;     REFERENCE/DOCKET NUMBER: P-LJ 2626
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (619) 535-9001
;     TELEFAX: (619) 535-8949
;   INFORMATION FOR SEQ ID NO: 23:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 1185 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-09-041-886-23
```

```
Query Match          42.5%; Score 161; DB 3; Length 1185;
Best Local Similarity 55.0%; Pred. No. 1.2e-10;
Matches 33; Conservative 0; Mismatches 5; Indels 22; Gaps 2;
```

```
Qy      7 VSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPGRLERP 66
          |||||      |||||
Db      476 VSTHHHHH-----QQQQQQQQQQQQQQHHGNSGPPP-PGAFPHP 513
```

RESULT 9

US-09-625-188-20

```
; Sequence 20, Application US/09625188
; Patent No. 6307037
; GENERAL INFORMATION:
;   APPLICANT: No. 6307037artis AG
```

```
; TITLE OF INVENTION: Fungal Target Genes and Methods
; FILE REFERENCE: PB/5-31285P1
; CURRENT APPLICATION NUMBER: US/09/625,188
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Ashbya gossypii
US-09-625-188-20
```

```
Query Match          41.8%; Score 158.5; DB 4; Length 729;
Best Local Similarity 45.5%; Pred. No. 1.3e-10;
Matches 40; Conservative 2; Mismatches 11; Indels 35; Gaps 4;
```

```
Qy          3 PRGSVSTHHH--HHQQQQQQQQQ---QQQQQQQQQQ----- 33
|:          ||  ||||| ||||  ||||  |||||
Db          448 PQSQALQHHLHHQQQQLQQQHHLQQQQHQQQQQSLSQQPQQQSQSQSAHSQQHQQQ 507

Qy          34 -QQQQQQQQQQQQQQQQQQHHGNSGPPEFP 60
| ||||| ||| |||  ||: |
Db          508 HQQQQQQQPQQQQPQQH-----PPQQP 530
```

RESULT 10

US-07-814-964-13

```
; Sequence 13, Application US/07814964
; Patent No. 5359047
```

; GENERAL INFORMATION:

```
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellett, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
```

```
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
```

; CORRESPONDENCE ADDRESS:

```
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
```

; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

; CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/07/814,964
; FILING DATE: 19911226
; CLASSIFICATION: 435
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: fractional yeast SSRP (fySSRP) (predicted)
US-07-814-964-13

```

```

Query Match          40.9%; Score 155; DB 1; Length 542;
Best Local Similarity 75.6%; Pred. No. 2.5e-10;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

Qy      11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH 51
        || | || | | |||| | || | | |||| | || | |
Db      231 HHQQQMQQQLQLQQQQQLQQQQQLQQQHQLQQQQQLQQQH 271

```

```

RESULT 11
US-08-258-442-13
; Sequence 13, Application US/08258442
; Patent No. 5670621
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellett, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/258,442
;   FILING DATE:
;   CLASSIFICATION:  530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  US 07/539,906
;   FILING DATE:  18-JUN-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Granahan, Patricia
;   REGISTRATION NUMBER:  32,227
;   REFERENCE/DOCKET NUMBER:  MIT-4787AAA
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  617-861-6240
;   TELEFAX:  617-861-9540
;   INFORMATION FOR SEQ ID NO:  13:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  542 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
;   ORIGINAL SOURCE:
;   ORGANISM:  Saccharomyces cerevisiae
;   IMMEDIATE SOURCE:
;   CLONE:  fractional yeast SSRP (fySSRP) (predicted)
US-08-258-442-13

```

```

Query Match          40.9%;  Score 155;  DB 1;  Length 542;
Best Local Similarity  75.6%;  Pred. No. 2.5e-10;
Matches  31;  Conservative  0;  Mismatches  10;  Indels  0;  Gaps  0;

```

```

Qy      11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH 51
        || | || | | |||| | |||| | || | |||| | ||||
Db      231 HHQQQMQQQLQLQQQQQLQQQQQLQQQHQLQQQQQLQQQH 271

```

# RESULT 12

US-08-328-809-8

```

; Sequence 8, Application US/08328809
; Patent No. 5705334

```

## GENERAL INFORMATION:

```

;   APPLICANT:  Lippard, Stephen J.
;   APPLICANT:  Essigmann, John M.
;   APPLICANT:  Donahue, Brian A.
;   APPLICANT:  Toney, Jeffrey H.
;   APPLICANT:  Bruhn, Suzanne L.
;   APPLICANT:  Pil, Pieter M.
;   APPLICANT:  Brown, Steven
;   APPLICANT:  Kellett, Patti
;   TITLE OF INVENTION:  Uses For DNA Structure-Specific
;   TITLE OF INVENTION:  Recognition Proteins
;   NUMBER OF SEQUENCES:  8
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  Patent Administrator, Testa, Hurwitz & Thibeault
;   STREET:  53 State Street
;   CITY:  Boston

```



```

; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,809
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fenton, Gillian M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-248-7000
; TELEFAX: 617-248-7100
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: fractional yeast SSRP (fySSRP) (predicted)
US-08-328-809-8

```

```

Query Match          40.9%; Score 155; DB 1; Length 542;
Best Local Similarity 75.6%; Pred. No. 2.5e-10;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

Qy      11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH 51
        || | ||| | |||| |||| || | |||| ||||
Db      231 HHQQQMQQQLQLQQQQQLQQQQQLQQQHQLQQQQQLQQQH 271

```

# RESULT 13

US-08-866-840-8

```

; Sequence 8, Application US/08866840
; Patent No. 6475791
; GENERAL INFORMATION:
; APPLICANT: Lippard, Stephen J.
; APPLICANT: Essigmann, John M.
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellett, Patti
; TITLE OF INVENTION: Uses For DNA Structure-Specific
; TITLE OF INVENTION: Recognition Proteins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: Patent Administrator, Testa, Hurwitz & Thibault  
 ; STREET: 53 State Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/866,840  
 ; FILING DATE: 02-JUN-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fenton, Gillian M.  
 ; REGISTRATION NUMBER: 36,508  
 ; REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-248-7000  
 ; TELEFAX: 617-248-7100  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 542 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Saccharomyces cerevisiae  
 ; IMMEDIATE SOURCE:  
 ; CLONE: fractional yeast SSRP (fySSRP) (predicted)  
 US-08-866-840-8

Query Match 40.9%; Score 155; DB 4; Length 542;  
 Best Local Similarity 75.6%; Pred. No. 2.5e-10;  
 Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH 51  
 || | ||| | |||| |||| | | |||| ||||  
 Db 231 HHQQQMQQQLQLQQQQQLQQQQQLQQQHQLQQQQQLQQQHH 271

# RESULT 14

PCT-US92-11107-13

; Sequence 13, Application PC/TUS9211107

## ; GENERAL INFORMATION:

; APPLICANT: Donahue, Brian A.  
 ; APPLICANT: Toney, Jeffrey H.  
 ; APPLICANT: Bruhn, Suzanne L.  
 ; APPLICANT: Pil, Pieter M.  
 ; APPLICANT: Brown, Steven  
 ; APPLICANT: Kellett, Patti  
 ; APPLICANT: Essigmann, John M.  
 ; APPLICANT: Lippard, Stephen J.  
 ; TITLE OF INVENTION: DNA Structure Specific Recognition  
 ; TITLE OF INVENTION: Protein and Uses Therefor

```

;   NUMBER OF SEQUENCES: 13
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET: 2 Militia Drive
;     CITY: Lexington
;     STATE: MA
;     COUNTRY: USA
;     ZIP: 02173
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US92/11107
;     FILING DATE: 19921218
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/539,906
;     FILING DATE: 18-JUN-1990
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Granahan, Patricia
;     REGISTRATION NUMBER: 32,227
;     REFERENCE/DOCKET NUMBER: MIT-4787AAA
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-861-6240
;     TELEFAX: 617-861-9540
;   INFORMATION FOR SEQ ID NO: 13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 542 amino acids
;       TYPE: AMINO ACID
;       TOPOLOGY: linear
;     MOLECULE TYPE: peptide
;     ORIGINAL SOURCE:
;       ORGANISM: Saccharomyces cerevisiae
;     IMMEDIATE SOURCE:
;       CLONE: fractional yeast SSRP (fySSRP) (predicted)
PCT-US92-11107-13

```

```

Query Match          40.9%;  Score 155;  DB 5;  Length 542;
Best Local Similarity 75.6%;  Pred. No. 2.5e-10;
Matches 31;  Conservative 0;  Mismatches 10;  Indels 0;  Gaps 0;

```

```

Qy      11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH 51
        || | ||| | |||| | |||| | || | |||| | ||||
Db      231 HHQQQMQQQLQLQQQQQLQQQQQLQQQHQLQQQQQLQQQH 271

```

```

RESULT 15
US-08-185-432-19
; Sequence 19, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
;   APPLICANT: Artavanis-Tsakonas, Spyridon
;   APPLICANT: Busseau, Isabelle
;   APPLICANT: Diederich, Robert J.
;   APPLICANT: Xu, Tian

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:36:59 ; Search time 11.5 Seconds  
(without alignments)  
577.149 Million cell updates/sec

Title: US-09-620-955B-9  
Perfect score: 379  
Sequence: 1 LVPRGSVSTHHHHHQQQQQQ.....HHGNSGPPEFPGRLERPHRD 69

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	%		Query			ID	Description
	Score	Match	Length	DB			
1	190.5	50.3	1154	2	S69206		regulator protein
2	189	49.9	905	1	RGBYS5		regulatory protein
3	188	49.6	1905	2	T18267		multidrug resistan
4	187	49.3	930	2	T08588		hypothetical prote
5	185	48.8	356	2	S31574		hypothetical prote
6	185	48.8	646	2	D82493		conserved hypothet
7	182	48.0	1090	2	A41696		regulatory protein
8	182	48.0	1761	2	T13675		hypothetical prote
9	181	47.8	1457	2	T14577		protein kinase Yak
10	180.5	47.6	758	2	S54522		hypothetical prote
11	179	47.2	339	1	TWHU2D		transcription init
12	179	47.2	1180	2	S69205		stripe a/b protein
13	179	47.2	1969	2	T08875		histidine kinase h

14	179	47.2	2150	2	S71629	sensory transducti
15	178.5	47.1	1023	2	T13068	CLOCK protein - fr
16	178.5	47.1	1027	2	T13071	CLOCK protein - fr
17	176.5	46.6	139	2	A26892	Mopa box protein -
18	173	45.6	644	2	S39356	transcription fact
19	173	45.6	708	2	A53185	G-box-binding fact
20	172	45.4	1010	2	T13167	Lola-like protein
21	171.5	45.3	320	2	E22364	alpha/beta-gliadin
22	171.5	45.3	1424	2	T03851	thyroid hormone re
23	169	44.6	770	2	T51024	related to C2H2 zi
24	169	44.6	873	2	B53225	ecdysone-induced p
25	168	44.3	613	2	S27770	hypothetical prote
26	168	44.3	1655	2	T13998	gene mastermind pr
27	167	44.1	816	2	S46268	ataxin-1 - human
28	166.5	43.9	1015	2	T13062	CLOCK protein - fr
29	166	43.8	700	2	S09699	bib protein - frui
30	166	43.8	853	2	T46347	hypothetical prote
31	165	43.5	1390	2	T14004	trfA protein - sli
32	164.5	43.4	648	1	JQ1150	protein kinase (EC
33	163	43.0	966	2	S25365	CYC8 protein - yea
34	161.5	42.6	482	2	B36901	chorion transcript
35	161.5	42.6	510	2	A36901	chorion transcript
36	161.5	42.6	514	2	C36901	chorion transcript
37	161	42.5	539	2	S57972	hypothetical prote
38	161	42.5	1184	2	G01763	atrophin-1 - human
39	161	42.5	1184	2	S50832	atrophin-1 - human
40	160.5	42.3	319	2	C22364	alpha/beta-gliadin
41	160.5	42.3	1085	2	S66149	gene pipsqueak pro
42	160.5	42.3	1893	2	A56158	eye development pr
43	160	42.2	1012	2	I53172	RAE-28 - mouse
44	160	42.2	2038	2	A43742	female sterile hom
45	159	42.0	354	2	S65687	(A+T)-stretch-bind

#### ALIGNMENTS

##### RESULT 1

S69206

regulator protein white collar 1 - *Neurospora crassa*

C;Species: *Neurospora crassa*

C;Date: 21-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 11-Jan-2002

C;Accession: S69206

R;Ballario, P.; Vittorioso, P.; Magrelli, A.; Talora, C.; Cabibbo, A.; Macino, G.

EMBO J. 15, 1650-1657, 1996

A;Title: White collar-1, a central regulator of blue light responses in *Neurospora*, is a zinc finger protein.

A;Reference number: S69206; MUID:96203083; PMID:8612589

A;Accession: S69206

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1154 <BAL>

A;Cross-references: EMBL:X94300; NID:gl279576; PID:gl480115

C;Genetics:

A;Introns: 967/3

C;Superfamily: GATA-type zinc finger homology

C;Keywords: zinc finger

F;932-991/Domain: GATA-type zinc finger homology <GZF>

Query Match 50.3%; Score 190.5; DB 2; Length 1154;  
Best Local Similarity 75.5%; Pred. No. 1.6e-10;  
Matches 40; Conservative 1; Mismatches 5; Indels 7; Gaps 2;

```
Qy      12 HHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ---HHGNSG---PP 57
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      20 HQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHQHQQQQKTNQHRNAGMMNTTP 72
```

## RESULT 2

RGBYS5

regulatory protein SNF5 - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: protein YBR2036; protein YBR289w

C;Species: *Saccharomyces cerevisiae*

C;Date: 30-Sep-1991 #sequence\_revision 09-Sep-1994 #text\_change 21-Jul-2000

C;Accession: S44551; S46171; A36375; S12067; S39145

R;Holmstrom, K.; Brandt, T.; Kallesoe, T.

Yeast 10(Suppl.A), S47-S62, 1994

A;Title: The sequence of a 32420 bp segment located on the right arm of chromosome II from *Saccharomyces cerevisiae*.

A;Reference number: S44537; MUID:94378722; PMID:8091861

A;Accession: S44551

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-905 <HOL>

A;Cross-references: EMBL:X76053; NID:g600025; PIDN:CAA53652.1; PID:g429134

R;Brandt, T.; Christiansen, C.; Holmstroem, K.; Kallesoe, T.

submitted to the Protein Sequence Database, August 1994

A;Reference number: S46157

A;Accession: S46171

A;Molecule type: DNA

A;Residues: 1-905 <BRA>

A;Cross-references: EMBL:Z36158; NID:g536741; PIDN:CAA85254.1; PID:g536742;

GSPDB:GN00002; MIPS:YBR289w

R;Laurent, B.C.; Treitel, M.A.; Carlson, M.

Mol. Cell. Biol. 10, 5616-5625, 1990

A;Title: The SNF5 protein of *Saccharomyces cerevisiae* is a glutamine- and proline-rich transcriptional activator that affects expression of a broad spectrum of genes.

A;Reference number: A36375; MUID:91042489; PMID:2233708

A;Accession: A36375

A;Molecule type: DNA

A;Residues: 1-563, 'D', 565-905 <LAU>

A;Cross-references: GB:M36482; NID:g172637; PIDN:AAA35062.1; PID:g172638

C;Genetics:

A;Gene: SGD:SNF5; MIPS:YBR289w

A;Cross-references: SGD:S0000493; MIPS:YBR289w

A;Map position: 2R

C;Superfamily: regulatory protein SNF5

C;Keywords: nucleus; transcription regulation

F;31-324/Region: glutamine/proline-rich

F;435-683/Region: acidic

F;714-882/Region: proline-rich

Query Match 49.9%; Score 189; DB 1; Length 905;  
Best Local Similarity 94.9%; Pred. No. 1.8e-10;  
Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 HQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHG 52  
|||||  
Db 231 HQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQG 269

#### RESULT 3

T18267

multidrug resistance protein - slime mold (*Dictyostelium discoideum*)

C;Species: *Dictyostelium discoideum*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T18267

R;Shaulsky, G.; Kuspa, A.; Loomis, W.F.

submitted to the EMBL Data Library, January 1995

A;Description: An MDR transporter/serine protease gene is required for prestalk specialization in *Dictyostelium*.

A;Reference number: Z18850

A;Accession: T18267

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1905 <SHA>

A;Cross-references: EMBL:U20432; NID:g664839; PID:g664840; PIDN:AAA62212.1

C;Genetics:

A;Gene: tagB

Query Match 49.6%; Score 188; DB 2; Length 1905;  
Best Local Similarity 86.0%; Pred. No. 4.4e-10;  
Matches 37; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 15 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPP 57  
|||||:|||||  
Db 1823 QQQQQEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQNDQPP 1865

#### RESULT 4

T08588

hypothetical protein L23H3.30 - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999

C;Accession: T08588

R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.

submitted to the Protein Sequence Database, May 1999

A;Reference number: Z16098

A;Accession: T08588

A;Molecule type: DNA

A;Residues: 1-930 <BEV>

A;Cross-references: EMBL:AL050398; GSPDB:GN00062; ATSP:L23H3.30

A;Experimental source: cultivar Columbia; BAC clone L23H3

C;Genetics:

A;Gene: ATSP:L23H3.30

A;Map position: 4

A;Introns: 11/2; 51/1; 87/3; 249/3; 278/2; 304/3; 330/1; 346/3; 449/3; 523/3;  
605/3; 645/1; 681/3; 723/3; 775/3; 814/3; 883/3



Query Match 49.3%; Score 187; DB 2; Length 930;  
Best Local Similarity 64.4%; Pred. No. 2.8e-10;  
Matches 38; Conservative 2; Mismatches 17; Indels 2; Gaps 1;

Qy 11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQ--QQQQQQQQQQQHHGNSGPPEFPGRLERPH 67  
| | | | | | | | | | | | | | | | | | | | | | : | | : |  
Db 139 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQHQNQPPSQQQQQQSTPQHQQQPTPQQQPQRRDGS 197

RESULT 5

S31574

hypothetical protein 2 - Mediterranean fruit fly

C;Species: Ceratitis capitata (Mediterranean fruit fly)

C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Sep-1997

C;Accession: S31574

R;Siden-Kiamos, I.; Favia, G.; Artiaco, D.; Saccone, G.; Furia, M.; Polito, L.C.; Louis, C.

submitted to the EMBL Data Library, December 1992

A;Description: Opa-like repeats in the genome of the Medfly Ceratitis capitata.

A;Reference number: S31573

A;Accession: S31574

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-356 <SID>

A;Cross-references: EMBL:X70053; NID:g5976; PID:g5977

Query Match 48.8%; Score 185; DB 2; Length 356;  
Best Local Similarity 51.9%; Pred. No. 1.8e-10;  
Matches 41; Conservative 2; Mismatches 14; Indels 22; Gaps 1;

Qy 11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQ-----Q 48  
| | | | | | | | | | | | | | | | | | :  
Db 137 HHAHQHMQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHTKEKLSERKMSVCKKEESSK 196

Qy 49 QHHGNSGPPEFPGRLERPH 67

: | | | | | | |

Db 197 RGAGNSNGQNFNSRTEDAH 215

RESULT 6

D82493

conserved hypothetical protein VCA0171 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C;Species: Vibrio cholerae

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C;Accession: D82493

R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Umayam, L.A.; Gill, S.R.; Nelson, K.E.; Read, T.D.; Tettelin, H.; Richardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; McDonald, L.; Utterback, T.; Fleishmann, R.D.; Nierman, W.C.; White, O.; Salzberg, S.L.; Smith, H.O.; Colwell, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: D82493  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-646 <HEI>  
 A;Cross-references: GB:AE004357; GB:AE003853; NID:g9657547; PIDN:AAF96084.1;  
 GSPDB:GN00127; TIGR:VCA0171  
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C;Genetics:  
 A;Gene: VCA0171  
 A;Map position: 2

Query Match 48.8%; Score 185; DB 2; Length 646;  
 Best Local Similarity 90.2%; Pred. No. 3.1e-10;  
 Matches 37; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 15 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSG 55  
 ||| :||  
 Db 454 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQDSSSG 494

# RESULT 7

A41696  
 regulatory protein nit-4 - Neurospora crassa  
 C;Species: Neurospora crassa  
 C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 24-Sep-1999  
 C;Accession: A41696; S37629; S20033  
 R;Yuan, G.F.; Fu, Y.H.; Marzluf, G.A.  
 Mol. Cell. Biol. 11, 5735-5745, 1991  
 A;Title: nit-4, a pathway-specific regulatory gene of Neurospora crassa, encodes a protein with a putative binuclear zinc DNA-binding domain.  
 A;Reference number: A41696; MUID:92017855; PMID:1840634  
 A;Accession: A41696  
 A;Molecule type: DNA  
 A;Residues: 1-1090 <YUA>  
 A;Cross-references: GB:M80368  
 R;Yuan, G.F.; Fu, Y.H.; Marzluf, G.A.  
 submitted to the EMBL Data Library, December 1991  
 A;Description: nit-4, a pathway-specific regulatory gene of Neurospora crassa, encodes a protein with a putative binuclear zinc DNA-binding domain.  
 A;Reference number: S37629  
 A;Accession: S37629  
 A;Molecule type: DNA  
 A;Residues: 1-98,'P',99-466,'S',468-1090 <YU2>  
 A;Cross-references: EMBL:M80368; NID:gl68848; PIDN:AAA33602.1; PID:gl68849  
 C;Genetics:  
 A;Gene: nit-4  
 A;Introns: 529/2  
 C;Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluster homology  
 C;Keywords: DNA binding; nucleus; transcription regulation; zinc finger  
 F;48-86/Domain: GAL4 zinc binuclear cluster homology <GAL4>

Query Match 48.0%; Score 182; DB 2; Length 1090;  
 Best Local Similarity 61.8%; Pred. No. 9.6e-10;  
 Matches 42; Conservative 4; Mismatches 2; Indels 20; Gaps 3;

Qy 1 LVPRGSV-----STHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ-- 49

```

          | |||::          ||          |:|||||:|||||||||||||||||||||
Db      971 LAPRGNIGGGGGGGGGST----GQRQQQQQRQQQQQQQQQQQQQQQQQQQQQQQQQEA 1026

Qy      50 -----HHG 52
          |||
Db      1027 NMFAYHHG 1034

```

RESULT 8

T13675  
hypothetical protein EG0002.3 - fruit fly (*Drosophila melanogaster*)  
C;Species: *Drosophila melanogaster*  
C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C;Accession: T13675  
R;Bolshakov, V.; Borkova, D.; Minana, B.; Kafatos, F.  
submitted to the EMBL Data Library, September 1998  
A;Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.  
A;Reference number: Z17698  
A;Accession: T13675  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1761 <BOL>  
A;Cross-references: EMBL:AL031130; NID:e1316407; PID:e1316410; PIDN:CAA20016.1  
C;Genetics:  
A;Cross-references: FlyBase:FBgn0025376  
A;Introns: 143/3; 237/3; 280/3  
A;Note: EG:EG0002.3

```

Query Match          48.0%; Score 182; DB 2; Length 1761;
Best Local Similarity 76.0%; Pred. No. 1.5e-09;
Matches 38; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

```

```

Qy      3 PRGSVSTHHHHHQ--QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQOH 50
          | |: : | |||||||||||||||||||||||||||||||||||
Db      1474 PAGATADMQRYVQRMQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQH 1523

```

RESULT 9

T14577  
protein kinase YakA (EC 2.7.1.-) - slime mold (*Dictyostelium discoideum*)  
C;Species: *Dictyostelium discoideum*  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C;Accession: T14577  
R;Kuspa, A.; Lu, S.; Souza, G.M.  
submitted to the EMBL Data Library, January 1998  
A;Description: YakA, a protein kinase required for the growth to development transition in *Dictyostelium*.  
A;Reference number: Z18146  
A;Accession: T14577  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-1457 <KUS>  
A;Cross-references: EMBL:AF045453; NID:g2854116; PID:g2854117; PIDN:AAC02554.1  
C;Genetics:  
A;Gene: yakA  
C;Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protein kinase

Query Match 47.8%; Score 181; DB 2; Length 1457;  
 Best Local Similarity 59.7%; Pred. No. 1.5e-09;  
 Matches 40; Conservative 3; Mismatches 10; Indels 14; Gaps 2;

Qy 13 HHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH-----GNSGPPEFPG 61  
 : ||||| :  
 Db 880 YQQQLQYQQQFQTLQDLNIEGEKPPPIYP- 938

Qy 62 RLERPHR 68  
 |||  
 Db 939 --NSPHR 943

# RESULT 10

S54522  
 hypothetical protein YMR164c - yeast (*Saccharomyces cerevisiae*)  
 N;Alternate names: hypothetical protein YM8520.13c  
 C;Species: *Saccharomyces cerevisiae*  
 C;Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 29-Oct-1999  
 C;Accession: S54522; S54609  
 R;Hunt, S.; Bowman, S.  
 submitted to the EMBL Data Library, May 1995  
 A;Reference number: S54510  
 A;Accession: S54522  
 A;Molecule type: DNA  
 A;Residues: 1-758 <HUN>  
 A;Cross-references: GB:Z49705; EMBL:Z49700; NID:g825556; PIDN:CAA89800.1;  
 PID:g825569; EMBL:Z49705; MIPS:YMR164c  
 A;Experimental source: strain AB972  
 C;Genetics:  
 A;Gene: SGD:MSS11  
 A;Cross-references: SGD:S0004774; MIPS:YMR164c  
 A;Map position: 13R

Query Match 47.6%; Score 180.5; DB 2; Length 758;  
 Best Local Similarity 61.5%; Pred. No. 9.5e-10;  
 Matches 40; Conservative 0; Mismatches 16; Indels 9; Gaps 2;

Qy 10 HHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGP-----PEFPGR 62  
 | ||| :  
 Db 287 HQPQHQPQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHQQQQQTPYPPIVNPQMVPHIPS- 345

Qy 63 LERPH 67  
 ||  
 Db 346 -ENSH 349

# RESULT 11

TWHU2D  
 transcription initiation factor IID - human  
 N;Alternate names: TATA-binding protein  
 C;Species: *Homo sapiens* (man)  
 C;Date: 20-Jul-1990 #sequence\_revision 19-May-1995 #text\_change 18-Feb-2000  
 C;Accession: A34830; A34831; S10944; I60128  
 R;Peterson, M.G.; Tanese, N.; Pugh, B.F.; Tjian, R.  
 Science 248, 1625-1630, 1990

A;Title: Functional domains and upstream activation properties of cloned human TATA binding protein.  
A;Reference number: A34830; MUID:90302006; PMID:2363050  
A;Accession: A34830  
A;Molecule type: mRNA  
A;Residues: 1-339 <PET>  
A;Cross-references: GB:M55654; NID:g339491; PIDN:AAA36731.1; PID:g339492  
R;Kao, C.C.; Lieberman, P.M.; Schmidt, M.C.; Zhou, Q.; Pei, R.; Berk, A.J. Science 248, 1646-1649, 1990  
A;Title: Cloning of a transcriptionally active human TATA binding factor.  
A;Reference number: A34831; MUID:90302010; PMID:2194289  
A;Accession: A34831  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-17,'N',19-186,'R',188-339 <KAO>  
R;Hoffmann, A.; Sinn, E.; Yamamoto, T.; Wang, J.; Roy, A.; Horikoshi, M.; Roeder, R.G. Nature 346, 387-390, 1990  
A;Title: Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID).  
A;Reference number: S10944; MUID:90326195; PMID:2374612  
A;Accession: S10944  
A;Molecule type: mRNA  
A;Residues: 1-91,96-339 <HOF>  
A;Cross-references: EMBL:X54993; NID:g37065; PIDN:CAA38736.1; PID:g37066  
R;Kao, C.; Lieberman, P.; Schmidt, M.; Zhou, Q.; Pei, R.; Berk, A.J. Science 248, 1626, 1990  
A;Title: Cloning of the human TATA binding factor: Expression of a transcriptionally active TFIID protein.  
A;Reference number: I60128  
A;Accession: I60128  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-186,'R',188-299,'MIKPR',300-339 <RES>  
A;Cross-references: GB:M34960; NID:g339493; PID:g339494  
C;Genetics:  
A;Gene: GDB:TBP; GTF2D1  
A;Cross-references: GDB:138768; OMIM:600075  
A;Map position: 6q27-6q27  
C;Superfamily: human transcription initiation factor IID  
C;Keywords: alternative splicing; DNA binding; nucleus; transcription initiation  
F;55-95/Region: glutamine-rich

Query Match 47.2%; Score 179; DB 1; Length 339;  
Best Local Similarity 81.8%; Pred. No. 6.4e-10;  
Matches 36; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 6 SVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 49  
|: ||||||||||||||||||||||||||||||||||  
Db 50 SILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 93

RESULT 12

S69205

stripe a/b protein - fruit fly (Drosophila melanogaster)

C;Species: Drosophila melanogaster

C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jul-2000

C;Accession: S69205  
R;Frommer, G.; Vorbrueggen, G.; Pasca, G.; Jaeckle, H.; Volk, T.  
EMBO J. 15, 1642-1649, 1996  
A;Title: Epidermal egr-like zinc finger protein of Drosophila participates in  
myotube guidance.  
A;Reference number: S69205; MUID:96203082; PMID:8612588  
A;Accession: S69205  
A;Status: preliminary; nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-1180 <FRO>  
A;Cross-references: EMBL:U42403; NID:g1147788; PIDN:AAB02355.1; PID:g1147789  
C;Keywords: alternative splicing

## RESULT 13

```
Query Match          47.2%;   Score 179;   DB 2;   Length 1969;
Best Local Similarity 65.0%;   Pred. No. 3.1e-09;
Matches    39;   Conservative    1;   Mismatches     4;   Indels     16;   Gaps      1;

Qy           6 SVSTHHHHH-----QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 49
               |:| | |                               |||||
Db          1694 SISDPHTSHLKGSSSHINQOIPSTIOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO 1753
```

## RESULT 14

S71629  
 sensory transduction histidine kinase dhkA - slime mold (Dictyostelium discoideum)  
 C;Species: Dictyostelium discoideum  
 C;Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 24-Sep-1998  
 C;Accession: S71629  
 R;Wang, N.; Shaulsky, G.; Escalante, R.; Loomis, W.F.  
 EMBO J. 15, 3890-3898, 1996  
 A;Title: A two-component histidine kinase gene that functions in Dictyostelium development.  
 A;Reference number: S71629; MUID:96324397; PMID:8670894  
 A;Accession: S71629  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-2150 <WAN>  
 A;Cross-references: EMBL:U42597  
 A;Experimental source: strain Ax4  
 C;Genetics:  
 A;Gene: dhkA  
 A;Map position: 6  
 C;Superfamily: response regulator homology  
 C;Keywords: autophosphorylation; phosphoprotein; phosphotransferase; two-component regulatory system  
 F;2027-2142/Domain: response regulator homology <RRH>  
 F;2076/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 47.2%; Score 179; DB 2; Length 2150;  
 Best Local Similarity 97.2%; Pred. No. 3.4e-09;  
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 16 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH 51  
 ||||| |||||  
 Db 33 QQQQLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH 68

RESULT 15  
 T13068  
 CLOCK protein - fruit fly (Drosophila melanogaster)  
 C;Species: Drosophila melanogaster  
 C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C;Accession: T13068  
 R;Darlington, T.K.; Wager-Smith, K.; Ceriani, M.F.; Staknis, D.; Gekakis, N.; Steeves, T.D.L.; Weitz, C.J.; Takahashi, J.S.; Kay, S.A.  
 Science 280, 1599-1603, 1998  
 A;Title: Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim.  
 A;Reference number: Z17599; MUID:98279147; PMID:9616122  
 A;Accession: T13068  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-1023 <DAR>  
 A;Cross-references: EMBL:AF067207; NID:g3192866; PID:g3192867; PIDN:AAD10630.1  
 C;Genetics:  
 A;Cross-references: FlyBase:FBgn0023076  
 C;Function:  
 A;Description: required for circadian behavioral rhythms

Query Match 47.1%; Score 178.5; DB 2; Length 1023;  
Best Local Similarity 60.3%; Pred. No. 1.9e-09;  
Matches 38; Conservative 2; Mismatches 4; Indels 19; Gaps 1;

```
Qy      6 SVSTHHH-----HQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 46
      :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      770 NLHTQHQNLLQQQHQSLSQLQQHTQQQHQQQQQQQQQQQQQQQQQQQQQQ 829

Qy      47 QQQ 49
      | |
Db      830 QLQ 832
```

Search completed: March 12, 2004, 15:41:45  
Job time : 12.5 secs



OM protein - protein search, using sw model

Run on: March 12, 2004, 15:39:10 ; Search time 27.3971 Seconds  
(without alignments)  
531.793 Million cell updates/sec

Title: US-09-620-955B-9  
Perfect score: 379  
Sequence: 1 LVPRGSVSTHHHHHQQQQQQ.....HHGNSGPPEFPGRLERPHRD 69

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

8

Result	Query					
No.	Score	Match	Length	DB	ID	Description

1	192	50.7	171	14	US-10-077-584-4	Sequence 4, Appli
2	188.5	49.7	910	10	US-09-086-436-31	Sequence 31, Appl
3	180.5	47.6	758	9	US-09-801-368-224	Sequence 224, App
4	179	47.2	338	9	US-09-933-638A-12	Sequence 12, Appl
5	179	47.2	339	15	US-10-116-275-184	Sequence 184, App
6	179	47.2	371	9	US-09-849-243-16	Sequence 16, Appl
7	179	47.2	2150	13	US-10-135-322-17	Sequence 17, Appl
8	178	47.0	1955	14	US-10-293-504-3	Sequence 3, Appli
9	176	46.4	97	9	US-09-864-761-35499	Sequence 35499, A
10	176	46.4	467	9	US-09-416-384A-7	Sequence 7, Appli
11	175	46.2	1138	14	US-10-074-475-194	Sequence 194, App
12	171.5	45.3	1420	14	US-10-379-616-4	Sequence 4, Appli
13	167	44.1	326	14	US-10-029-386-32987	Sequence 32987, A
14	167	44.1	816	14	US-10-207-706-3	Sequence 3, Appli
15	166	43.8	1070	9	US-09-735-367B-6	Sequence 6, Appli
16	166	43.8	2005	9	US-09-735-367B-3	Sequence 3, Appli
17	166	43.8	2063	9	US-09-735-367B-2	Sequence 2, Appli
18	165	43.5	80	14	US-10-177-725-14	Sequence 14, Appl
19	165	43.5	406	15	US-10-369-493-3147	Sequence 3147, Ap
20	163	43.0	966	9	US-09-801-368-372	Sequence 372, App
21	156.5	41.3	1572	15	US-10-116-275-179	Sequence 179, App
22	156	41.2	623	15	US-10-464-939-12	Sequence 12, Appl
23	156	41.2	780	9	US-09-770-689A-5	Sequence 5, Appli
24	154	40.6	264	13	US-10-029-180-30	Sequence 30, Appl
25	154	40.6	944	13	US-10-029-180-26	Sequence 26, Appl
26	153	40.4	4952	15	US-10-051-874-56	Sequence 56, Appl
27	153	40.4	5008	15	US-10-051-874-166	Sequence 166, App
28	153	40.4	5159	15	US-10-085-198-112	Sequence 112, App
29	153	40.4	5262	15	US-10-051-874-165	Sequence 165, App
30	153	40.4	5262	15	US-10-051-874-167	Sequence 167, App
31	152.5	40.2	702	14	US-10-161-051-18	Sequence 18, Appl
32	150.5	39.7	170	9	US-09-864-761-42294	Sequence 42294, A
33	150.5	39.7	1221	14	US-10-270-333-60	Sequence 60, Appl
34	150	39.6	429	9	US-09-987-107-34	Sequence 34, Appl
35	149.5	39.4	314	14	US-10-317-832-13	Sequence 13, Appl
36	149.5	39.4	905	15	US-10-369-493-5635	Sequence 5635, Ap
37	149.5	39.4	905	15	US-10-369-493-5636	Sequence 5636, Ap
38	149	39.3	72	10	US-09-820-843A-14	Sequence 14, Appl
39	148.5	39.2	398	15	US-10-374-780A-2358	Sequence 2358, Ap
40	147.5	38.9	736	9	US-09-922-364A-47	Sequence 47, Appl
41	147.5	38.9	736	9	US-09-254-590-47	Sequence 47, Appl
42	147.5	38.9	736	13	US-10-115-695-47	Sequence 47, Appl
43	147.5	38.9	736	14	US-10-116-561-47	Sequence 47, Appl
44	147.5	38.9	736	14	US-10-115-671-47	Sequence 47, Appl
45	147.5	38.9	736	14	US-10-115-415-47	Sequence 47, Appl

#### ALIGNMENTS

##### RESULT 1

US-10-077-584-4

; Sequence 4, Application US/10077584

; Publication No. US20030073610A1

; GENERAL INFORMATION:

; APPLICANT: LINDQUIST, SUSAN



RESULT 3

US-09-801-368-224

; Sequence 224, Application US/09801368

; Patent No. US20020128250A1

; GENERAL INFORMATION:

; APPLICANT: Busby, Robert

; APPLICANT: Cali, Brian

; APPLICANT: Hecht, Peter

; APPLICANT: Holtzman, Doug

; APPLICANT: Madden, Kevin

; APPLICANT: Maxon, Mary

; APPLICANT: Milne, Todd

; APPLICANT: No. US20020128250A1man, Thea

; APPLICANT: Royer, John

; APPLICANT: Salama, Sofie

; APPLICANT: Sherman, Amir

; APPLICANT: Silva, Jeff

; APPLICANT: Summers, Eric

; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

; FILE REFERENCE: 109272.147

; CURRENT APPLICATION NUMBER: US/09/801,368

; CURRENT FILING DATE: 2001-03-07

; PRIOR APPLICATION NUMBER: US 09/487,558

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: US 60/160,587

; PRIOR FILING DATE: 1999-10-20

; NUMBER OF SEQ ID NOS: 440

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 224

; LENGTH: 758

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-801-368-224

Query Match 47.6%; Score 180.5; DB 9; Length 758;

Best Local Similarity 61.5%; Pred. No. 3e-09;

Matches 40; Conservative 0; Mismatches 16; Indels 9; Gaps 2;

Qy 10 HHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGP-----PEFPGR 62  
| || |||||

Db 287 HQPQHQPQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHQQQQTPYPPIVNPQMVPHIPS- 345

Qy 63 LERPH 67  
| |

Db 346 -ENSH 349

RESULT 4

US-09-933-638A-12

; Sequence 12, Application US/09933638A

; Patent No. US20020160952A1

; GENERAL INFORMATION:

; APPLICANT: Kazantsev, Aleksey G.

; APPLICANT: Thompson, Leslie M.

; APPLICANT: Housman, David E.

```
; TITLE OF INVENTION: INHIBITION OF PROTEIN-PROTEIN INTERACTION
; FILE REFERENCE: 01997-289001
; CURRENT APPLICATION NUMBER: US/09/933,638A
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: US 60/226,502
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-933-638A-12
```

```
Query Match          47.2%; Score 179; DB 9; Length 338;
Best Local Similarity 81.8%; Pred. No. 2e-09;
Matches 36; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
```

```
Qy      6 SVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 49
      |:      |||||||||||||||||||||||||||||||||||
Db      50 SILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 93
```

# RESULT 5

US-10-116-275-184

```
; Sequence 184, Application US/10116275
; Publication No. US20030211476A1
```

## ; GENERAL INFORMATION:

```
; APPLICANT: Elan Pharmaceutical Technology
; APPLICANT: O'Mahony, Daniel J.
; APPLICANT: Brayden, David
; APPLICANT: Byrne, Daragh
; APPLICANT: Lambkin, Imelda
; APPLICANT: Higgins, Lisa
```

```
; TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and
Methods and
```

```
; TITLE OF INVENTION: Compositions Targeting Peyer's Patches and M Cell
Receptors
```

```
; FILE REFERENCE: E1067/20087
; CURRENT APPLICATION NUMBER: US/10/116,275
; CURRENT FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 349
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 184
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-275-184
```

```
Query Match          47.2%; Score 179; DB 15; Length 339;
Best Local Similarity 81.8%; Pred. No. 2e-09;
Matches 36; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
```

```
Qy      6 SVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 49
      |:      |||||||||||||||||||||||||||||||||||
Db      50 SILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 93
```

US-09-849-243-16

US-09-849-243-16

Query Match 47.2%; Score 179; DB 9; Length 371;  
Best Local Similarity 81.8%; Pred. No. 2.2e-09;  
Matches 36; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy           6 SVSTHHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ     49  
             |:           ||||||||||||||||||||||||||||||||||  
Db          82 SILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ     125

US-10-135-322-17

; Sequence 17, Application US/10135322

; Publication No. US20020173017A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BENFEY, PN  
 ; APPLICANT: HELARIUTTA, Y  
 ; APPLICANT: MAHONEN, AP  
 ; APPLICANT: BONKE, AWM  
 ; APPLICANT: KAUPPINEN, L  
 ; APPLICANT: RIIKONEN, M  
 ; TITLE OF INVENTION: WOODEN LEG GENE, PROMOTER AND USES THEREOF  
 ; FILE REFERENCE: 5914-086-999  
 ; CURRENT APPLICATION NUMBER: US/10/135,322  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: 60/253,739  
 ; PRIOR FILING DATE: 2000-11-29  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 17  
 ; LENGTH: 2150  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 US-10-135-322-17

Query Match 47.2%; Score 179; DB 13; Length 2150;  
 Best Local Similarity 97.2%; Pred. No. 1.1e-08;  
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 16 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH 51  
 |||| ||||||||||||||||||||||||||||||||  
 Db 33 QQQQLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHH 68

RESULT 8

US-10-293-504-3  
 ; Sequence 3, Application US/10293504  
 ; Publication No. US20030110520A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Universitdegli studi di Roma La Sapienza  
 ; APPLICANT: Macino, Giuseppe  
 ; APPLICANT: Cogoni, Carlo  
 ; TITLE OF INVENTION: Isolation and characterization of a N. crassa silencing  
 ; TITLE OF INVENTION: gene and uses therof  
 ; FILE REFERENCE: PC  
 ; CURRENT APPLICATION NUMBER: US/10/293,504  
 ; CURRENT FILING DATE: 2002-11-13  
 ; PRIOR APPLICATION NUMBER: US/09/857,091  
 ; PRIOR FILING DATE: 2001-05-31  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 1955  
 ; TYPE: PRT  
 ; ORGANISM: Neurospora crassa  
 US-10-293-504-3

Query Match 47.0%; Score 178; DB 14; Length 1955;  
 Best Local Similarity 66.7%; Pred. No. 1.2e-08;  
 Matches 36; Conservative 5; Mismatches 9; Indels 4; Gaps 1;

Qy 5 GSVSTHHHHHQQQQQQQQQQQQQQQ-----QQQQQQQQQQQQQQQQQQQQHHGNS 54  
 || : | ||| |||||::|: | ||||| ||||| ||||| ||||| :|  
 Db 42 GSSTFDHEQHQQHQQQQQKQRQSQSEARQQQQQQQQQQQQQQQQQQQAQHHHS 95

RESULT 9

US-09-864-761-35499

; Sequence 35499, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES  
 USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29





; ORGANISM: mus musculus  
US-09-416-384A-7

RESULT 11

; Sequence 194, Application US/10074475

; GENERAL INFORMATION:

; APPLICANT: Salceda, Susana

; APPLICANT: Macina, Roberto

; APPLICANT: Hu, Ping

; APPLICANT: Recipon, Herve

; APPLICANT: Karra, Kalpana

; APPLICANT: Cafferkey, Robert

; APPLICANT: Sun, Yongming

; APPLICANT: Liu, Chenghua

; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific

; TITLE OF INVENTION: Genes and Proteins

FILE REFERENCE: DEX-0313

; CURRENT APPLICATION NUMBER: US/10/074,475

; CURRENT FILING DATE: 2002-02-13

; PRIOR APPLICATION NUMBER: 60/268,292

; PRIOR FILING DATE: 2001-02-13

; NUMBER OF SEQ ID NOS: 295

```
; SOFTWARE: PatentIn version 3.1
```

; SEO ID NO 194

; LENGTH: 1138

```
; TYPE: PRT
```

; ORGANISM: Homo sapien

US-10-074-475-194

Query Match 46.2%; Score 175; DB 14; Length 1138;  
Best Local Similarity 56.3%; Pred. No. 1.4e-08;  
Matches 40; Conservative 2; Mismatches 17; Indels 12; Gaps 2;

Qy            2 VPRGSVSTHHHHHQQQQQQQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNS-- 54  
             || :| | ||| | |:||||||| | | | | | | | | | | | | | | | |  
Db          457 VPSSDMSPAELKOMAAOQOORAKLMOOKOQOQOQOQOQOQOQOQOQOQOQOHSNOTS 516

Qy            55 -----GPPEFP    60  
               ||| |  
Db            517 NWSPLGPPSSP    527

RESULT 12

; Sequence 4, Application US/10379616

; Publication No. US20030153047A1



Matches 42; Conservative 4; Mismatches 20; Indels 18; Gaps 2;

```

Qy      1 LVPRGSVS-THHHHHQXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----- 45
      |  ||:| | | :||||||||||| | |||||||||||||||
Db      181 LANMGSLSQTPGHKAEQXXXXXXXXXQHQXXXXXXXXXXXXXXXXXHLSRAPGLITPGSP 240

Qy      46 ---QQQHHGNSGPPEFPGRLERP 66
      || |: | |: || |
Db      241 PPAQQNQYVHISSSPQNTGRTASP 264

```

RESULT 14

US-10-207-706-3

```

; Sequence 3, Application US/10207706
; Publication No. US20030143681A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M.
; TITLE OF INVENTION: Human Ataxin-1-Like Polypeptide IMX97018
; FILE REFERENCE: 3138-A
; CURRENT APPLICATION NUMBER: US/10/207,706
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 816
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-207-706-3

```

Query Match 44.1%; Score 167; DB 14; Length 816;  
 Best Local Similarity 50.0%; Pred. No. 5.9e-08;  
 Matches 42; Conservative 4; Mismatches 20; Indels 18; Gaps 2;

```

Qy      1 LVPRGSVS-THHHHHQXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----- 45
      |  ||:| | | :||||||||||| | |||||||||||||||
Db      181 LANMGSLSQTPGHKAEQXXXXXXXXXQHQXXXXXXXXXXXXXXXXXHLSRAPGLITPGSP 240

Qy      46 ---QQQHHGNSGPPEFPGRLERP 66
      || |: | |: || |
Db      241 PPAQQNQYVHISSSPQNTGRTASP 264

```

RESULT 15

US-09-735-367B-6

```

; Sequence 6, Application US/09735367B
; Patent No. US20020151477A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Caira, Francoise
; APPLICANT: Antonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05

```

; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1070  
; TYPE: PRT  
; ORGANISM: mammal  
US-09-735-367B-6

Query Match 43.8%; Score 166; DB 9; Length 1070;  
Best Local Similarity 50.0%; Pred. No. 9.4e-08;  
Matches 42; Conservative 4; Mismatches 16; Indels 22; Gaps 3;

```
Qy      3 PRGSVSTHHHHHQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQQ----- 42
      | ||:: || |               ||||| ||||| ||||| ||||| ||||| |
Db      232 PSGSLAPPHHPMQPVSVNRMNPANFPQLQQQQQQQQQQQQQQQQQQQQQQQQQLQARPP 291

Qy      43 QQQQQQQHHGNSGPPEFPGRLEP 66
      || |||| | | | : |
Db      292 QQHQQQQPQGIR--PQFTAPTQVP 313
```

Search completed: March 12, 2004, 15:44:13  
Job time : 27.3971 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:34:19 ; Search time 33.8235 Seconds  
(without alignments)  
643.657 Million cell updates/sec

Title: US-09-620-955B-9  
Perfect score: 379  
Sequence: 1 LVPRGSVSTHHHHHQQQQQ.....HHGNSGPPEFPGRLERPHRD 69

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

%  
Result Query  
No. Score Match Length DB ID Description  
-----

1	218.5	57.7	92	6	Q9GM66	Q9gm66	equus cabal
2	210.5	55.5	313	6	O97927	O97927	pan paniscu
3	210	55.4	1461	5	Q86AZ9	Q86az9	dictyosteli
4	210	55.4	1485	5	Q8MMV4	Q8mmv4	dictyosteli
5	207.5	54.7	2472	5	Q8MXN1	Q8mxn1	dictyosteli
6	202	53.3	108	6	O18905	O18905	canis famil
7	200	52.8	556	5	O76940	O76940	drosophila
8	199.5	52.6	3469	5	Q9U4I2	Q9u4i2	drosophila
9	199.5	52.6	3604	5	Q9VYK0	Q9vyk0	drosophila
10	197	52.0	2294	5	Q9VUB7	Q9vub7	drosophila
11	194	51.2	1191	4	Q86V38	Q86v38	homo sapien
12	193.5	51.1	680	5	Q86AM9	Q86am9	dictyosteli
13	191	50.4	652	5	Q8T2S4	Q8t2s4	dictyosteli
14	191	50.4	1521	5	Q86AB8	Q86ab8	dictyosteli
15	191	50.4	1811	5	Q8IJD3	Q8ijd3	plasmodium
16	191	50.4	2123	5	Q9U9S7	Q9u9s7	dictyosteli
17	190	50.1	71	5	Q8MP18	Q8mpl8	dictyosteli
18	189	49.9	816	5	Q86HD8	Q86hd8	dictyosteli
19	189	49.9	2230	5	Q86A58	Q86a58	dictyosteli
20	188	49.6	739	11	Q7TPU6	Q7tpu6	mus musculu
21	188	49.6	856	5	Q8T151	Q8t151	dictyosteli
22	188	49.6	3770	5	Q869R6	Q869r6	dictyosteli
23	187.5	49.5	726	5	Q86H70	Q86h70	dictyosteli
24	187	49.3	149	4	Q8NFT3	Q8nft3	homo sapien
25	187	49.3	215	6	Q8MJ84	Q8mj84	pongo pygma
26	187	49.3	217	4	Q8NOW2	Q8n0w2	homo sapien
27	187	49.3	218	6	Q8MHX3	Q8mhx3	pan troglod
28	187	49.3	222	4	Q8Nfq4	Q8nfq4	homo sapien
29	187	49.3	365	4	Q8Nfq1	Q8nfq1	homo sapien
30	187	49.3	415	4	Q8Nfq3	Q8nfq3	homo sapien
31	187	49.3	431	4	Q8N6B6	Q8n6b6	homo sapien
32	187	49.3	456	4	Q8N6B5	Q8n6b5	homo sapien
33	187	49.3	522	13	O42323	O42323	coturnix co
34	187	49.3	586	11	Q8BQ27	Q8bq27	mus musculu
35	187	49.3	713	6	Q8MJ98	Q8mj98	pongo pygma
36	187	49.3	713	6	Q8MJ99	Q8mj99	gorilla gor
37	187	49.3	714	6	Q8MJ97	Q8mj97	macaca mula
38	187	49.3	714	6	Q8HYZ9	Q8hyz9	pongo pygma
39	187	49.3	714	11	Q8R441	Q8r441	mus musculu
40	187	49.3	714	11	Q8C4F0	Q8c4f0	mus musculu
41	187	49.3	716	6	Q8MJA0	Q8mja0	pan troglod
42	187	49.3	716	6	Q8HZ00	Q8hz00	pan paniscu
43	187	49.3	740	4	Q8IZE0	Q8ize0	homo sapien
44	187	49.3	1037	5	Q867Z5	Q867z5	drosophila
45	187	49.3	1502	5	Q8IS10	Q8is10	dictyosteli

#### ALIGNMENTS

##### RESULT 1

Q9GM66

ID Q9GM66 PRELIMINARY; PRT; 92 AA.

AC Q9GM66;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Atrophin-1 (Fragment).  
 GN DRPLA.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tozaki T.;  
 RT "equine dentatorubral-pallidoluysian atrophy (DRPLA) gene and  
 RT microsatellite locus; TKY30.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB048336; BAB13349.1; -.  
 DR InterPro; IPR002951; Atrophin.  
 DR Pfam; PF03154; Atrophin-1; 2.  
 FT NON\_TER 1 1  
 FT NON\_TER 92 92  
 SQ SEQUENCE 92 AA; 10280 MW; AE6A07C0B8B4ED1E CRC64;

Query Match 57.7%; Score 218.5; DB 6; Length 92;  
 Best Local Similarity 74.1%; Pred. No. 2.1e-17;  
 Matches 43; Conservative 1; Mismatches 5; Indels 9; Gaps 2;

Qy 9 THHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPGRLERP 66  
 ||||| |  
 Db 10 THHHHH-----QQQQQQQQQQQQQQQQQQQQQQQQHHGSSGPPP-PGAYPHP 58

# RESULT 2

O97927

ID O97927 PRELIMINARY; PRT; 313 AA.  
 AC O97927;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Atrophin-1 (Fragment).  
 GN DRPLA.  
 OS Pan paniscus (Pygmy chimpanzee) (Bonobo).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OX NCBI\_TaxID=9597;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gangeswaran R., Chana H.S., Santibanez-Koref M.F., Hancock J.M.;  
 RT "Evolution of three triplet expansion genes.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AJ133270; CAB37923.1; -.  
 DR InterPro; IPR002951; Atrophin.  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR Pfam; PF03154; Atrophin-1; 1.  
 DR PRINTS; PR01222; ATROPHIN.  
 DR PRINTS; PR01217; PRICHEXTENSIN.  
 FT NON\_TER 1 1  
 FT NON\_TER 313 313  
 SQ SEQUENCE 313 AA; 31862 MW; 7DA5D62F192AE822 CRC64;

Query Match 55.5%; Score 210.5; DB 6; Length 313;







RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AC117080; AAM45327.2; -.  
 DR InterPro; IPR002423; Cpn60/TCP-1.  
 DR Pfam; PF00118; cpn60\_TCP1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 2472 AA; 278497 MW; 30CCCF7157D4008A7 CRC64;

```

QY      1  LVPRGSV-----STHHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ  49
          |  |||:          |  |||||||  |||||||  |||||||  |||||||  |||||||
Db     199  LSPRGSILRSNSQQHQHQHQHQQQQQQQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ  252

```

018905

Query Match 53.3%; Score 202; DB 6; Length 108;  
Best Local Similarity 69.0%; Pred. No. 1.8e-15;  
Matches 40; Conservative 2; Mismatches 10; Indels 6; Gaps 2;

Qy           9 THHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQHGHGNSGPPEFPGRLERP 66  
             | | | | |     | | | | |     | : | | | | | | | | | | | | : | | | | |  
Db          23 THHHHH-----QQQQQQPPPPQSQRPPQQQQQQQQQQQQQQQQQQHGHGSSGPPP-PGAYPHP 74

076940

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE GAGA factor class A-isoform.  
 GN TRL OR TRITHORAX-LIKE.  
 OS Drosophila virilis (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7244;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lintermann K.G., Roth G.E., King-Jones K., Korge G., Lehmann M.;  
 RT "Comparison of the GAGA factor genes of Drosophila melanogaster and  
 RT Drosophila virilis reveals high conservation of GAGA factor structure  
 RT beyond the BTB/POZ and DNA-binding domains.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AJ005174; CAA06415.1; -.  
 DR HSSP; Q08605; 1YUI.  
 DR FlyBase; FBgn0025647; Dvir\Trl.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005515; F:protein binding; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR001137; Glyco\_hydro\_11.  
 DR InterPro; IPR000408; Reg\_chrom\_condens.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00651; BTB; 1.  
 DR Pfam; PF00096; zf-C2H2; 1.  
 DR SMART; SM00225; BTB; 1.  
 DR SMART; SM00355; Znf\_C2H2; 1.  
 DR PROSITE; PS50097; BTB; 1.  
 DR PROSITE; PS00777; GLYCOSYL\_HYDROL\_F11\_2; 1.  
 DR PROSITE; PS00626; RCC1\_2; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 556 AA; 60117 MW; 581AF0C95DF888CE CRC64;

Query Match 52.8%; Score 200; DB 5; Length 556;  
 Best Local Similarity 72.2%; Pred. No. 1.4e-14;  
 Matches 39; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 1 LVPRGSVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNS 54  
 :|: : || |||||  
 Db 444 VLPQQQLQQQHQTTPQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHLNTS 497

# RESULT 8

Q9U4I2

ID Q9U4I2 PRELIMINARY; PRT; 3469 AA.  
 AC Q9U4I2;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE SANT domain protein SMRTER.  
 GN SMR OR SMRTER OR CG4013.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99417957; PubMed=10488333;  
 RA Tsai C.-C., Kao H.-Y., Yao T.-P., McKeown M., Evans R.M.;  
 RT "SMRTER, a Drosophila nuclear receptor coregulator, reveals that EcR-  
 RT mediated repression is critical for development.";  
 RL Mol. Cell 4:175-186(1999).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.  
 DR EMBL; AF175223; AAD52614.1; -.  
 DR FlyBase; FBgn0024308; Smr.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR002086; Aldehyde\_dehydr.  
 DR InterPro; IPR001005; Myb\_DNA\_binding.  
 DR Pfam; PF00249; myb\_DNA-binding; 1.  
 DR SMART; SM00717; SANT; 1.  
 DR PROSITE; PS00687; ALDEHYDE\_DEHYDR\_GLU; 1.  
 KW DNA-binding; Nuclear protein.  
 SQ SEQUENCE 3469 AA; 364115 MW; 6284E14C5C247CD9 CRC64;

Query Match 52.6%; Score 199.5; DB 5; Length 3469;  
 Best Local Similarity 78.4%; Pred. No. 9.3e-14;  
 Matches 40; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

Qy 11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPG 61  
 |||||||||||||||||||||:|||| |||| | || ||  
 Db 88 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQKQQQHHMQQQQQQQPLS-PPHPPG 137

# RESULT 9

Q9VYK0

ID Q9VYK0 PRELIMINARY; PRT; 3604 AA.  
 AC Q9VYK0;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CG4013 protein.  
 GN SMR OR CG4013.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanenvong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003490; AAF48195.2; -.  
 DR FlyBase; FBgn0024308; Smr.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR002086; Aldehyde\_dehydr.  
 DR InterPro; IPR001005; Myb\_DNA\_binding.  
 DR Pfam; PF00249; myb\_DNA-binding; 1.  
 DR PROSITE; PS00687; ALDEHYDE DEHYDR GLU; 1.  
 SQ SEQUENCE 3604 AA; 378155 MW; B7563A180C1D546B CRC64;

Query Match 52.6%; Score 199.5; DB 5; Length 3604;  
 Best Local Similarity 78.4%; Pred. No. 9.7e-14;  
 Matches 40; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

Qy 11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPEFPG 61  
 ||||||||||||||||||||||||||||:|||| |||| | || ||  
 Db 216 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQKQQQHHMQQQQQQPLS-PPHPPG 265

# RESULT 10

Q9VUB7

ID Q9VUB7 PRELIMINARY; PRT; 2294 AA.  
 AC Q9VUB7;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE CG32133 protein.  
 GN CG32133 OR CG6532 OR CG8797.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of *Drosophila melanogaster* genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,



RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003536; AAF49771.2; -.  
 DR FlyBase; FBgn0052133; CG32133.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR InterPro; IPR001357; BRCT.  
 DR Pfam; PF00533; BRCT; 6.  
 DR SMART; SM00292; BRCT; 6.  
 DR PROSITE; PS50172; BRCT; 4.  
 SQ SEQUENCE 2294 AA; 262480 MW; 4A18D9B6C645CD17 CRC64;

Query Match 52.0%; Score 197; DB 5; Length 2294;  
 Best Local Similarity 92.7%; Pred. No. 1.2e-13;  
 Matches 38; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 10 HHHHHQXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXH 50  
 | | ||||||||||||||||||||||||||||||||||||  
 Db 217 HQHQMQXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXH 257

# RESULT 11

Q86V38

ID Q86V38 PRELIMINARY; PRT; 1191 AA.  
 AC Q86V38;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE DRPLA protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,



DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
SQ SEQUENCE 680 AA; 79759 MW; 6DB9FE3034BAC068 CRC64;

```

Qy      8  STHHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHGG-----NSGPPEFP 60
      |      |||||||||||||||||||||||||||||||||||      |      ||:
Db      558  SQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHGYQPPQQYNHQPQQHQ 617

Qy      61  GRLERPHR 68
      :  :  |:
Db      618  HOHOHOHQ 625

```

Q8T2S4

Query Match 50.4%; Score 191; DB 5; Length 652;  
Best Local Similarity 100.0%; Pred. No. 1.7e-13;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 14

```

ID      Q86AB8                PRELIMINARY;          PRT;   1521 AA.
AC      Q86AB8;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Similar to plasmodium falciparum. malaria antigen.
OS      Dictyostelium discoideum (Slime mold).
OC      Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX      NCBI_TaxID=44689;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AX4;
RX      MEDLINE=22092622; PubMed=12097910;
RA      Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA      Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA      Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT      "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL      Nature 418:79-85(2002).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AX4;
RA      Baumgart C.;
RL      Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AC115577; AAO51739.1; -.
DR      InterPro; IPR008938; ARM.
SQ      SEQUENCE      1521 AA;  171723 MW;  6FE1CD644CF2E80B CRC64;

```

Qy 7 VSTH-----HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNS 54  
:| | : |||||  
Db 845 ISIHNSGGIVNSOoooHNNNT 898

## Q81JD3

AC Q8IJD3;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF10\_0265.  
OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,  
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Nature 419:498-511(2002).  
 DR EMBL; AE014833; AAN35462.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1811 AA; 216655 MW; 8A25116576D5FED1 CRC64;

Query Match 50.4%; Score 191; DB 5; Length 1811;  
 Best Local Similarity 92.5%; Pred. No. 4.5e-13;  
 Matches 37; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 15 QQQHHGNS 54  
 ||||||||||||||||||||||||||||||||||||||||:|:  
 Db 532 QQHNNNN 571

Search completed: March 12, 2004, 15:40:58  
 Job time : 34.8235 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:22:04 ; Search time 8.11765 Seconds  
(without alignments)  
442.596 Million cell updates/sec

Title: US-09-620-955B-9  
Perfect score: 379  
Sequence: 1 LVPRGSVSTHHHHHQQQQQQ.....HHGNSGPPEFPGRLERPHRD 69

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query		DB	ID	Description
	Score	Match Length			
1	190.5	50.3	1167	1 WC1_NEUCR	Q01371 neurospora
2	190.5	50.3	1516	1 NCO2_XENLA	Q9w705 xenopus lae
3	189	49.9	905	1 SNF5_YEAST	P18480 saccharomyc
4	188.5	49.7	910	1 HCN1_MOUSE	O88704 mus musculu
5	188	49.6	1177	1 SP97_DICDI	Q95zg3 dictyosteli
6	188	49.6	1905	1 TAGB_DICDI	P54683 dictyosteli
7	187	49.3	714	1 FXP2_MOUSE	P58463 mus musculu
8	187	49.3	715	1 FXP2_HUMAN	O15409 homo sapien
9	187	49.3	716	1 FXP2_PANTR	Q8mja0 pan troglod
10	187	49.3	931	1 LUG_ARATH	Q9fuy2 arabidopsis
11	182	48.0	1090	1 NIT4_NEUCR	P28349 neurospora
12	181	47.8	705	1 FXP1_MOUSE	P58462 mus musculu
13	180.5	47.6	758	1 YM38_YEAST	Q03825 saccharomyc
14	180.5	47.6	2212	1 T230_HUMAN	Q93074 homo sapien
15	179	47.2	339	1 TBP_HUMAN	P20226 homo sapien
16	178.5	47.1	1023	1 CLOC_DROME	O61735 drosophila
17	174	45.9	2067	1 NCO6_MOUSE	Q9j119 m nuclear r

18	173	45.6	644	1	BTD_DROME	Q24266	drosophila
19	173	45.6	708	1	GBF_DICDI	P36417	dictyosteli
20	171.5	45.3	1424	1	NCO3_HUMAN	Q9y6q9	h nuclear r
21	169	44.6	527	1	RBF1_CANAL	Q00312	candida alb
22	169	44.6	1080	1	HDC_DROME	Q9n2m8	drosophila
23	167	44.1	816	1	ATX1_HUMAN	P54253	homo sapien
24	166	43.8	700	1	BIB_DROME	P23645	drosophila
25	166	43.8	1161	1	BM2K_HUMAN	Q9nsyl	homo sapien
26	166	43.8	2063	1	NCO6_HUMAN	Q14686	h nuclear r
27	164.5	43.4	648	1	KAPC_DICDI	P34099	dictyosteli
28	164	43.3	1138	1	BM2K_MOUSE	Q91z96	mus musculu
29	163	43.0	966	1	SSN6_YEAST	P14922	saccharomyc
30	162	42.7	3726	1	ABF1_MOUSE	Q61329	mus musculu
31	161.5	42.6	510	1	CF2_DROME	P20385	drosophila
32	161.5	42.6	514	1	CF23_DROME	Q01522	drosophila
33	161	42.5	1185	1	DRPL_HUMAN	P54259	homo sapien
34	160	42.2	1073	1	HR38_DROME	P49869	drosophila
35	160	42.2	2038	1	FSH_DROME	P13709	drosophila
36	159	42.0	910	1	HCN1_RAT	Q9jkb0	rattus norv
37	158	41.7	829	1	E74A_DROME	P20105	drosophila
38	158	41.7	883	1	E74B_DROME	P11536	drosophila
39	158	41.7	1012	1	PHC1_MOUSE	Q64028	mus musculu
40	157.5	41.6	615	1	CPO_DROME	Q01617	drosophila
41	157.5	41.6	1319	1	MN1_HUMAN	Q10571	homo sapien
42	157	41.4	395	1	SRY_MOUSE	Q05738	mus musculu
43	157	41.4	645	1	BRH2_DROME	Q24256	drosophila
44	156.5	41.3	1586	1	SN22_HUMAN	P51531	homo sapien
45	156	41.2	623	1	DSH_DROME	P51140	drosophila

#### ALIGNMENTS

##### RESULT 1

##### WC1\_NEUCR

ID WC1\_NEUCR STANDARD; PRT; 1167 AA.  
 AC Q01371;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE White collar 1 protein (WC1).  
 GN WC-1.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=74-OR23-1A / FGSC 987;  
 RX MEDLINE=96203083; PubMed=8612589;  
 RA Ballario P., Vittorioso P., Magrelli A., Talora C., Cabibbo A.,  
 RA Macino G.;  
 RT "White collar-1, a central regulator of blue light responses in  
 RT Neurospora, is a zinc finger protein."  
 RL EMBO J. 15:1650-1657(1996).  
 RN [2]  
 RP REVISIONS TO C-TERMINUS.

RA Ballario P.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR INVOLVED IN LIGHT  
 CC REGULATION. BINDS AND AFFECTS BLUE LIGHT REGULATION OF THE AL-3  
 CC GENE. WC1 AND WC2 PROTEINS INTERACT VIA HOMOLOGOUS PAS DOMAINS,  
 CC BIND TO PROMOTERS OF LIGHT REGULATED GENES SUCH AS FRQ, AND  
 CC ACTIVATE TRANSCRIPTION.  
 CC -!- SUBUNIT: HETERODIMER OF WC1 AND WC2 (POTENTIAL).  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- INDUCTION: By blue light.  
 CC -!- DOMAIN: The glutamine-rich domain might function in activating  
 CC gene expression.  
 CC -!- SIMILARITY: Contains 1 GATA-type zinc finger.  
 CC -!- SIMILARITY: Contains 3 PAS (PER-ARNT-SIM) dimerization domains.  
 CC -!- SIMILARITY: Contains 2 PAS-associated C-terminal (PAC) domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X94300; CAA63964.2; -.  
 DR HSSP; P17679; 1GNF.  
 DR TRANSFAC; T02819; -.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000014; PAS\_domain.  
 DR InterPro; IPR000679; Znf\_GATA.  
 DR Pfam; PF00320; GATA; 1.  
 DR Pfam; PF00785; PAC; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR SMART; SM00086; PAC; 2.  
 DR SMART; SM00091; PAS; 3.  
 DR SMART; SM00401; ZnF\_GATA; 1.  
 DR TIGRFAMs; TIGR00229; sensory\_box; 3.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
 DR PROSITE; PS50114; GATA\_ZN\_FINGER\_2; 1.  
 DR PROSITE; PS50112; PAS; 3.  
 KW Transcription regulation; Activator; DNA-binding; Zinc-finger;  
 KW Nuclear protein; Repeat.  
 FT DOMAIN 16 61 GLN-RICH.  
 FT DOMAIN 381 452 PAS 1.  
 FT DOMAIN 469 508 PAC 1.  
 FT DOMAIN 574 644 PAS 2.  
 FT DOMAIN 650 691 PAC 2.  
 FT DOMAIN 693 763 PAS 3.  
 FT ZN\_FING 934 959 GATA-TYPE.  
 FT DOMAIN 21 57 POLY-GLN.  
 FT DOMAIN 329 333 POLY-PRO.  
 SQ SEQUENCE 1167 AA; 127454 MW; 6489D04DAB50EE38 CRC64;

Query Match 50.3%; Score 190.5; DB 1; Length 1167;  
 Best Local Similarity 75.5%; Pred. No. 3.5e-09;  
 Matches 40; Conservative 1; Mismatches 5; Indels 7; Gaps 2;



Qy 12 HHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ---HHGNSG----PP 57  
| | | | | | | | | | | | | | | | | | | | | | | | | | | : | |  
Db 20 HQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHQQHQQQQKTNQHRNAGMMNTTP 72

## RESULT 2

```

NC02_XENLA
ID NCO2_XENLA STANDARD; PRT; 1516 AA.
AC Q9W705;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nuclear receptor coactivator 2 (NCoA-2) (Transcriptional intermediary
DE factor 2) (XTIF2).
GN NCOA2 OR TIF2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171035; PubMed=10704837;
RA de la Calle-Mustienes E., Gomez-Skarmeta J.L.;
RT "XTIF2, a Xenopus homologue of the human transcription intermediary
RT factor, is required for a nuclear receptor pathway that also
RT interacts with CBP to suppress Brachyury and XMyoD.";
RL Mech. Dev. 91:119-129(2000).
CC -!- FUNCTION: Transcriptional coactivator for steroid receptors and
CC nuclear receptors. Coactivator of the steroid binding domain
CC (AF-2) but not of the modulating N-terminal domain (AF-1).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: Expressed homogeneously during late blastula-
CC early gastrula stage and later becomes highly expressed in the
CC notochord.
CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC -!- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ243119; CAB45389.1; -.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR000014; PAS_domain.
DR InterPro; IPR008955; Src-1.
DR Pfam; PF00989; PAS; 1.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00091; PAS; 1.
DR PROSITE; PS50888; HLH; 1.
DR PROSITE; PS50112; PAS; 1.
KW Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 116 180 PAS.

```

FT DOMAIN 1237 1273 POLY-GLN.  
SQ SEQUENCE 1516 AA; 166156 MW; 09851C00AB439A4A CRC64;

Query Match 50.3%; Score 190.5; DB 1; Length 1516;  
Best Local Similarity 64.6%; Pred. No. 4.4e-09;  
Matches 42; Conservative 2; Mismatches 14; Indels 7; Gaps 1;

Qy 4 RGSVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQH-----HGNSGP 56  
| :| | | | | | | | | | | | | | | | | | | | | | | | | | : |  
Db 1228 REILSQHLRQKQLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHRAMMMRSQGLAMP 1287  
  
Qy 57 PEFPG 61  
| |  
Db 1288 PNMVG 1292

### RESULT 3

#### SNF5\_YEAST

ID SNF5\_YEAST STANDARD; PRT; 905 AA.  
AC P18480;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Transcription regulatory protein SNF5 (SWI/SNF complex component SNF5)  
DE (Transcription factor TYE4).  
GN SNF5 OR TYE4 OR SWI10 OR YBR289W OR YBR2036.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MCY;  
RX MEDLINE=91042489; PubMed=2233708;  
RA Laurent B.C., Treitel M.A., Carlson M.;  
RT "The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and  
RT proline-rich transcriptional activator that affects expression of a  
RT broad spectrum of genes.";  
RL Mol. Cell. Biol. 10:5616-5625(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RX MEDLINE=94378722; PubMed=8091861;  
RA Holmstroem K., Brandt T., Kallesoe T.;  
RT "The sequence of a 32,420 bp segment located on the right arm of  
RT chromosome II from Saccharomyces cerevisiae.";  
RL Yeast 10:S47-S62(1994).  
CC -!- FUNCTION: Involved in transcriptional activation. The SWI/SNF  
CC complex is required for the induced expression of a large number  
CC of genes. This complex alters chromatin structure to facilitate  
CC binding of gene-specific dedicated transcription factors.  
CC -!- SUBUNIT: Component of the SWI/SNF global transcription activator  
CC complex.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: Belongs to the SNF5 family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M36482; AAA35062.1; -.

DR EMBL; X76053; CAA53652.1; -.

DR EMBL; Z36158; CAA85254.1; -.

DR PIR; S44551; RGBYS5.

DR GermOnline; 138832; -.

DR SGD; S0000493; SNF5.

DR InterPro; IPR006939; SNF5.

DR Pfam; PF04855; SNF5; 1.

KW Transcription regulation; Activator; Nuclear protein.

FT	DOMAIN	31	270	GLN-RICH.
----	--------	----	-----	-----------

FT	DOMAIN	72	132	PRO-RICH.
----	--------	----	-----	-----------

FT	DOMAIN	272	324	PRO-RICH.
----	--------	-----	-----	-----------

FT	DOMAIN	489	588	ASP/GLU-RICH (ACIDIC).
----	--------	-----	-----	------------------------

FT	DOMAIN	714	882	PRO-RICH.
----	--------	-----	-----	-----------

FT      DOMAIN      755      798      ARG/LYS-RICH (BASIC).

FT CONFLICT 564 564 E -> D (IN REF. 1).

SQ SEQUENCE 905 AA; 102557 MW; A287B4A648DD1A35 CRC64;

Query Match 49.9%; Score 189; DB 1; Length 905;

Best Local Similarity 94.9%; Pred. No. 3.7e-09;

Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy            14 H00000000000000000000000000000000000000H0HG 52

[illegible]

Db            231 HQQG 269

## RESULT 4

HCN1 MOUSE

ID HCN1 MOUSE STANDARD; PRT: 910 AA.

AC 088704; 054899; 09D613;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated

DE channel 1 (Brain cyclic nucleotide gated channel 1) (BCNG-1)

DE (Hyperpolarization-activated cation channel 2) (HAC-2).

GN HCN1 OR BCNG1 OR HAC2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A., AND N-GLYCOSYLATION.

RC STRAIN=C57BL/6J; TISSUE=Brain;

RX MEDLINE=98070835; PubMed=9405696;

RA Santoro B., Grant S.G.N., Bartsch D., Kandel E.R.;

RT "Interactive cloning with the SH3 domain of N-src identifies a new

RT brain specific ion channel protein, with homology to eag and cyclic

RT nucleotide-gated channels.";

RL Proc. Natl. Acad. Sci. U.S.A. 94:14815-14820(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Brain;  
 RX MEDLINE=98295993; PubMed=9634236;  
 RA Ludwig A., Zong X., Jeglitsch M., Hofmann F., Biel M.;  
 RT "A family of hyperpolarization-activated cation channels.";  
 RL Nature 393:587-591(1998).  
 RN [3]  
 RP SEQUENCE OF 377-910 FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [4]  
 RP FUNCTION, AND REGULATION BY CAMP.  
 RX MEDLINE=98292171; PubMed=9630217;  
 RA Santoro B., Liu D.T., Yao H., Bartsch D., Kandel E.R.,  
 RA Siegelbaum S.A., Tibbs G.R.;  
 RT "Identification of a gene encoding a hyperpolarization-activated  
 RT pacemaker channel of brain.";  
 RL Cell 93:717-729(1998).  
 RN [5]  
 RP INTERACTION WITH KCNE2.  
 RX MEDLINE=21313430; PubMed=11420311;  
 RA Yu H., Wu J., Potapova I., Wymore R.T., Holmes B., Zuckerman J.,  
 RA Pan Z., Wang H., Shi W., Robinson R.B., El-Maghrabi M.R., Benjamin W.,  
 RA Dixon J.E., McKinnon D., Cohen I.S., Wymore R.;  
 RT "MinK-related peptide 1: A beta subunit for the HCN ion channel  
 RT subunit family enhances expression and speeds activation.";  
 RL Circ. Res. 88:E84-E87(2001).  
 RN [6]  
 RP REGULATION BY CAMP.  
 RX MEDLINE=21351681; PubMed=11459060;  
 RA Wainger B.J., DeGennaro M., Santoro B., Siegelbaum S.A., Tibbs G.R.;  
 RT "Molecular mechanism of cAMP modulation of HCN pacemaker channels.";  
 RL Nature 411:805-810(2001).  
 RN [7]  
 RP FUNCTION, AND TISSUE SPECIFICITY.

RX MEDLINE=21530492; PubMed=11675786;  
RA Stevens D.R., Seifert R., Bufe B., Mueller F., Kremmer E., Gauss R.,  
RA Meyerhof W., Kaupp U.B., Lindemann B.;  
RT "Hyperpolarization-activated channels HCN1 and HCN4 mediate responses  
RT to sour stimuli.";  
RL Nature 413:631-635(2001).

RN [8]  
RP INTERACTION WITH HCN2, AND MUTAGENESIS OF GLY-349; TYR-350 AND  
RP GLY-351.

RX MEDLINE=22083667; PubMed=12089064;  
RA Xue T., Marban E., Li R.A.;  
RT "Dominant-negative suppression of HCN1- and HCN2-encoded pacemaker  
RT currents by an engineered HCN1 construct: insights into  
RT structure-function relationships and multimerization.";  
RL Circ. Res. 90:1267-1273(2002).

RN [9]  
RP OLIGOMERIZATION VIA N-TERMINAL DOMAIN.

RX MEDLINE=22162449; PubMed=12034718;  
RA Proenza C., Tran N., Angoli D., Zahynacz K., Balcar P., Accili E.A.;  
RT "Different roles for the cyclic nucleotide binding domain and amino  
RT terminus in assembly and expression of hyperpolarization-activated,  
RT cyclic nucleotide-gated channels.";  
RL J. Biol. Chem. 277:29634-29642(2002).

RN [10]  
RP MUTAGENESIS OF CYS-303 AND CYS-318.

RX MEDLINE=22336443; PubMed=12351622;  
RA Xue T., Li R.A.;

RT "An external determinant in the S5-P linker of the pacemaker (HCN)  
RT channel identified by sulfhydryl modification.";  
RL J. Biol. Chem. 277:46233-46242(2002).

CC -!- FUNCTION: Hyperpolarization-activated ion channel exhibiting weak  
CC selectivity for potassium over sodium ions. Contributes to the  
CC native pacemaker currents in heart (If) and in neurons (Ih).  
CC Activated by cAMP, and at 10-100 times higher concentrations, also  
CC by cGMP. May mediate responses to sour stimuli.

CC -!- SUBUNIT: The potassium channel is probably composed of a homo- or  
CC heterotetrameric complex of pore-forming subunits. Heteromultimer  
CC with HCN2. Interacts with KCNE2. Interacts with the SH3 domain of  
CC CSK.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- TISSUE SPECIFICITY: Predominantly expressed in brain. Highly  
CC expressed in apical dendrites of pyramidal neurons in the cortex,  
CC in the layer corresponding to the stratum lacunosum-moleculare in  
CC the hippocampus and in axons of basket cells in the cerebellum.  
CC Expressed in a subset of elongated cells in taste buds.

CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is  
CC characterized by a series of positively charged amino acids at  
CC every third position.

CC -!- PTM: N-glycosylated.

CC -!- MISCELLANEOUS: Inhibited by extracellular cesium ions.

CC -!- SIMILARITY: Belongs to the potassium channel family. HCN  
CC subfamily.

CC -!- SIMILARITY: Contains 1 cyclic nucleotide-binding domain.

CC -!- CAUTION: Ref.3 sequence differs from that shown due to a  
CC frameshift in position 381.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF028737; AAC53518.1; -.  
 DR EMBL; AJ225123; CAA12407.1; -.  
 DR EMBL; AK014722; BAB29519.1; ALT\_FRAME.  
 DR MGD; MGI:1096392; Hcn1.  
 DR InterPro; IPR000595; cNMP\_binding.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR005820; M+channel\_nlg.  
 DR Pfam; PF00027; cNMP\_binding; 1.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR SMART; SM00100; cNMP; 1.  
 DR PROSITE; PS00888; CNMP\_BINDING\_1; 1.  
 DR PROSITE; PS00889; CNMP\_BINDING\_2; FALSE\_NEG.  
 DR PROSITE; PS50042; CNMP\_BINDING\_3; 1.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Sodium transport;  
 KW cAMP; cAMP-binding; Transmembrane; Glycoprotein; Sodium channel.  
 FT DOMAIN 1 135 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 136 156 SEGMENT S1 (POTENTIAL).  
 FT TRANSMEM 163 183 SEGMENT S2 (POTENTIAL).  
 FT DOMAIN 184 208 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 209 229 SEGMENT S3 (POTENTIAL).  
 FT TRANSMEM 238 258 SEGMENT S4 (POTENTIAL).  
 FT DOMAIN 259 289 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 290 310 SEGMENT S5 (POTENTIAL).  
 FT TRANSMEM 334 355 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
 FT TRANSMEM 361 381 SEGMENT S6 (POTENTIAL).  
 FT DOMAIN 382 910 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 78 129 INVOLVED IN SUBUNIT ASSEMBLY (BY  
 FT SIMILARITY).  
 FT NP\_BIND 464 581 CAMP.  
 FT DOMAIN 1 81 GLY-RICH.  
 FT DOMAIN 715 777 GLN-RICH.  
 FT DOMAIN 878 884 POLY-PRO.  
 FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT MUTAGEN 303 303 C->S: ABOLISHES CONDUCTIVITY.  
 FT MUTAGEN 318 318 C->S: ABOLISHES SENSITIVITY TO SULFHYDRIL  
 FT MODIFICATION.  
 FT MUTAGEN 349 349 G->A: ABOLISHES CONDUCTIVITY; WHEN  
 FT ASSOCIATED WITH A-350 AND A-351.  
 FT MUTAGEN 350 350 Y->A: ABOLISHES CONDUCTIVITY; WHEN  
 FT ASSOCIATED WITH A-349 AND A-351.  
 FT MUTAGEN 351 351 G->A: ABOLISHES CONDUCTIVITY; WHEN  
 FT ASSOCIATED WITH A-349 AND A-350.  
 FT CONFLICT 42 42 G -> R (IN REF. 1).  
 FT CONFLICT 394 394 R -> S (IN REF. 3).  
 SQ SEQUENCE 910 AA; 102432 MW; 56FD5F328DD972E9 CRC64;

Query Match 49.7%; Score 188.5; DB 1; Length 910;  
 Best Local Similarity 66.7%; Pred. No. 4.1e-09;



SQ SEQUENCE 1177 AA; 136812 MW; C45B848B016A94ED CRC64;

Query Match 49.6%; Score 188; DB 1; Length 1177;  
Best Local Similarity 79.6%; Pred. No. 5.7e-09;  
Matches 39; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 LVPRGSVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 49  
| || | | |||||  
Db 1001 LEPRFSYQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 1049

RESULT 6

TAGB\_DICDI

ID TAGB\_DICDI STANDARD; PRT; 1905 AA.

AC P54683;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Prestalk-specific protein tagB precursor (EC 3.4.21.-).

GN TAGB.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

OX NCBI\_TaxID=44689;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AX4;

RX MEDLINE=95262903; PubMed=7744252;

RA Shaulsky G., Kuspa A., Loomis W.F.;

RT "A multidrug resistance transporter/serine protease gene is required  
for prestalk specialization in Dictyostelium.";

RL Genes Dev. 9:1111-1122(1995).

CC -!- FUNCTION: Intercellular communication via tagB may mediate  
integration of cellular differentiation with morphogenesis.

CC -!- SIMILARITY: In the N-terminal section; belongs to peptidase family  
S8.

CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING  
TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.

CC -!- SIMILARITY: STRONG, TO TAGC.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; U20432; AAA62212.1; -.

DR PIR; T18267; T18267.

DR MEROPS; S08.UPW; -.

DR DictyBase; DDB0001964; tagB.

DR InterPro; IPR003593; AAA\_ATPase.

DR InterPro; IPR001140; ABC\_TM\_transpt.

DR InterPro; IPR003439; ABC\_transporter.

DR InterPro; IPR000209; Peptidase\_S8.

DR Pfam; PF00664; ABC\_membrane; 1.

DR Pfam; PF00005; ABC\_tran; 1.



DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS50929; ABC\_TM1F; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; FALSE\_NEG.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KW Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;  
 KW Signal.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 32 1905 PRESTALK-SPECIFIC PROTEIN TAGB.  
 FT DOMAIN 378 700 PROTEASE.  
 FT DOMAIN 1518 1756 ABC TRANSPORTER.  
 FT TRANSMEM 1011 1031 POTENTIAL.  
 FT TRANSMEM 1076 1096 POTENTIAL.  
 FT TRANSMEM 1121 1141 POTENTIAL.  
 FT TRANSMEM 1210 1230 POTENTIAL.  
 FT TRANSMEM 1309 1329 POTENTIAL.  
 FT TRANSMEM 1332 1352 POTENTIAL.  
 FT ACT\_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 432 432 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 695 695 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP\_BIND 1553 1560 ATP (POTENTIAL).  
 FT DOMAIN 63 67 POLY-GLN.  
 FT DOMAIN 95 104 POLY-ASN.  
 FT DOMAIN 107 134 POLY-ASN.  
 FT DOMAIN 311 321 POLY-SER.  
 FT DOMAIN 833 837 POLY-SER.  
 FT DOMAIN 838 844 POLY-GLY.  
 FT DOMAIN 871 876 POLY-LEU.  
 FT DOMAIN 1012 1015 POLY-ILE.  
 FT DOMAIN 1386 1389 POLY-GLU.  
 FT DOMAIN 1398 1404 POLY-GLY.  
 FT DOMAIN 1445 1450 POLY-ASN.  
 FT DOMAIN 1765 1779 POLY-ASN.  
 FT DOMAIN 1782 1785 POLY-SER.  
 FT DOMAIN 1807 1812 POLY-PRO.  
 FT DOMAIN 1813 1860 POLY-GLN.  
 FT DOMAIN 1872 1878 POLY-PRO.  
 FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 747 747 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1172 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1522 1522 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1658 1658 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1905 AA; 212518 MW; B8E223FA8B9AE13C CRC64;

Query Match 49.6%; Score 188; DB 1; Length 1905;  
 Best Local Similarity 86.0%; Pred. No. 8.6e-09;  
 Matches 37; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 15 QQQHHGNSGPP 57

Db 1823 |||||:||||| 1 11  
QQQQQEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQNDQPP 1865

RESULT 7

FXP2\_MOUSE

ID FXP2\_MOUSE STANDARD; PRT; 714 AA.  
AC P58463;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Forkhead box protein P2.  
GN FOXP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Lung;  
RX MEDLINE=21347947; PubMed=11358962;  
RA Shu W., Yang H., Zhang L., Lu M.M., Morrissey E.E.;  
RT "Characterization of a new subfamily of winged-helix/forkhead (Fox)  
RT genes that are expressed in the lung and act as transcriptional  
RT repressors.";  
RL J. Biol. Chem. 276:27488-27497(2001).  
CC -!- FUNCTION: Transcriptional repressor that play an important role in  
CC the specification and differentiation of lung epithelium. May play  
CC important roles in developing neural, gastrointestinal and  
CC cardiovascular tissues.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- TISSUE SPECIFICITY: Highest expression in lung. Lower expression  
CC in spleen, skeletal muscle, brain, kidney and small intestine.  
CC -!- DEVELOPMENTAL STAGE: Expressed in developing lung (only distal  
CC epithelium), neural, intestinal and cardiovascular tissues.  
CC -!- SIMILARITY: Contains 1 fork-head domain.  
CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF339106; AAK69651.1; -.  
DR MGD; MGI:2148705; Foxp2.  
DR GO; GO:0016564; F:transcriptional repressor activity; IDA.  
DR GO; GO:0016481; P:negative regulation of transcription; IDA.  
DR InterPro; IPR001766; TF\_Fork\_head.  
DR InterPro; IPR007087; ZnF\_C2H2.  
DR Pfam; PF00250; Fork\_head; 1.  
DR PRINTS; PR00053; FORKHEAD.  
DR ProDom; PD000425; TF\_Fork\_head; 1.  
DR SMART; SM00339; FH; 1.  
DR SMART; SM00355; ZnF\_C2H2; 1.

DR PROSITE; PS00657; FORK\_HEAD\_1; FALSE\_NEG.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; FALSE\_NEG.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; FALSE\_NEG.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein.  
 FT ZN\_FING 345 370 C2H2-TYPE.  
 FT DNA\_BIND 503 593 FORK-HEAD.  
 FT DOMAIN 53 56 POLY-GLN.  
 FT DOMAIN 123 126 POLY-GLN.  
 FT DOMAIN 131 136 POLY-GLN.  
 FT DOMAIN 152 191 POLY-GLN.  
 FT DOMAIN 200 208 POLY-GLN.  
 FT DOMAIN 222 230 POLY-GLN.  
 SQ SEQUENCE 714 AA; 79820 MW; BCDFB80E28398609 CRC64;

Query Match 49.3%; Score 187; DB 1; Length 714;  
 Best Local Similarity 97.4%; Pred. No. 4.5e-09;  
 Matches 37; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 15 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHG 52  
 ||||||||||||||||||||||||||||||||||||  
 Db 157 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHPG 194

# RESULT 8

## FXP2\_HUMAN

ID FXP2\_HUMAN STANDARD; PRT; 715 AA.  
 AC O15409; Q8NOW2;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Forkhead box protein P2 (CAG repeat protein 44) (Trinucleotide repeat-  
 DE containing gene 10 protein).  
 GN FOXP2 OR CAGH44 OR TNRC10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANT SPCH1 HIS-553.  
 RX MEDLINE=21470412; PubMed=11586359;  
 RA Lai C.S.L., Fisher S.E., Hurst J.A., Vargha-Khadem F., Monaco A.P.;  
 RT "A forkhead-domain gene is mutated in a severe speech and language  
 RT disorder.";  
 RL Nature 413:519-523(2001).  
 RN [2]  
 RP SEQUENCE OF 1-304 FROM N.A.  
 RC TISSUE=Brain cortex;  
 RX MEDLINE=97369492; PubMed=9225980;  
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,  
 RA Breschel T.S., Stine O.C., Callahan C., Mcinnis M.G., Ross C.A.;  
 RT "cDNAs with long CAG trinucleotide repeats from human brain.";  
 RL Hum. Genet. 100:114-122(1997).  
 RN [3]  
 RP SEQUENCE OF 1-86 FROM N.A.

RA Minx P., Hinds K., Sutterer C., Becker M., Ozersky P.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 113-329 FROM N.A.  
 RX MEDLINE=22179809; PubMed=12192408;  
 RA Enard W., Przeworski M., Fisher S.E., Lai C.S.L., Wiebe V., Kitano T.,  
 RA Monaco A.P., Paabo S.;  
 RT "Molecular evolution of FOXP2, a gene involved in speech and  
 RT language.";  
 RL Nature 418:869-872(2002).  
 CC -!- FUNCTION: Transcriptional repressor that plays an important role  
 CC in the specification and differentiation of lung epithelium. May  
 CC play important roles in developing neural, gastrointestinal and  
 CC cardiovascular tissues. Involved in neural mechanisms mediating  
 CC the development of speech and language.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1; Synonyms=I;  
 CC IsoId=O15409-1; Sequence=Displayed;  
 CC Name=2; Synonyms=II;  
 CC IsoId=O15409-3; Sequence=Not described;  
 CC Name=3; Synonyms=III, IV;  
 CC IsoId=O15409-2; Sequence=VSP\_001558;  
 CC -!- TISSUE SPECIFICITY: Expressed at high levels in embryonic and  
 CC adult lung.  
 CC -!- DISEASE: Defects in FOXP2 are the cause of speech-language  
 CC disorder 1 (SPCH1) [MIM:602081]; also known as autosomal dominant  
 CC speech and language disorder with orofacial dyspraxia. Affected  
 CC individuals have a severe impairment in the selection and  
 CC sequencing of fine orofacial movements, which are necessary for  
 CC articulation. They also show deficits in several facets of  
 CC language processing (such as the ability to break up words into  
 CC their constituent phoneme) and grammatical skills.  
 CC -!- DISEASE: Disruption of FOXP2 by a chromosomal translocation  
 CC t(5;7)(q22;q31.2) is the cause of severe speech and language  
 CC impairment.  
 CC -!- SIMILARITY: Contains 1 fork-head domain.  
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF337817; AAL10762.1; -.  
 DR EMBL; U80741; AAB91439.1; -.  
 DR EMBL; AC003992; -; NOT ANNOTATED\_CDS.  
 DR EMBL; AF515031; AAN03389.1; -.  
 DR EMBL; AF515032; AAN03390.1; -.  
 DR EMBL; AF515033; AAN03391.1; -.  
 DR EMBL; AF515034; AAN03392.1; -.  
 DR EMBL; AF515035; AAN03393.1; -.  
 DR EMBL; AF515036; AAN03394.1; -.

```

DR   EMBL; AF515037; AAN03395.1; -.
DR   EMBL; AF515038; AAN03396.1; -.
DR   EMBL; AF515039; AAN03397.1; -.
DR   EMBL; AF515040; AAN03398.1; -.
DR   EMBL; AF515041; AAN03399.1; -.
DR   EMBL; AF515042; AAN03400.1; -.
DR   EMBL; AF515043; AAN03401.1; -.
DR   EMBL; AF515044; AAN03402.1; -.
DR   EMBL; AF515045; AAN03403.1; -.
DR   EMBL; AF515046; AAN03404.1; -.
DR   EMBL; AF515047; AAN03405.1; -.
DR   EMBL; AF515048; AAN03406.1; -.
DR   EMBL; AF515049; AAN03407.1; -.
DR   EMBL; AF515050; AAN03408.1; -.
DR   Genew; HGNC:13875; FOXP2.
DR   MIM; 605317; -.
DR   MIM; 602081; -.
DR   InterPro; IPR001766; TF_Fork_head.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM00339; FH; 1.
DR   SMART; SM00355; Znf_C2H2; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR   PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Chromosomal translocation; Disease mutation;
KW   Alternative splicing.
FT   ZN_FING      346      371      C2H2-TYPE.
FT   DNA_BIND     504      594      FORK-HEAD.
FT   DOMAIN       53       56      POLY-GLN.
FT   DOMAIN      123      126      POLY-GLN.
FT   DOMAIN      131      136      POLY-GLN.
FT   DOMAIN      152      191      POLY-GLN.
FT   DOMAIN      200      209      POLY-GLN.
FT   DOMAIN      223      231      POLY-GLN.
FT   VARSPLIC     1       92      Missing (in isoform 3).
FT                                     /FTId=VSP_001558.
FT   VARIANT      553      553      R -> H (in SPCH1).
FT                                     /FTId=VAR_012278.
FT   CONFLICT     134      134      Q -> H (IN REF. 2).
FT   CONFLICT     290      304      DLTTNNSSTTSNT -> EEFPVQGPAAVCAGL (IN
FT                                     REF. 2).
SQ   SEQUENCE     715 AA;  79919 MW;  4F9FBDB6D90516E0 CRC64;

Query Match          49.3%;  Score 187;  DB 1;  Length 715;
Best Local Similarity 97.4%;  Pred. No. 4.5e-09;
Matches    37;  Conservative    0;  Mismatches    1;  Indels      0;  Gaps      0;

Qy      15  QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHG 52
          |||||||||||||||||||||||||||||||||
Db      157  QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHPG 194

```

RESULT 9

FXP2\_PANTR

ID FXP2\_PANTR STANDARD; PRT; 716 AA.  
AC Q8MJA0; Q8MHX3;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Forkhead box protein P2.  
GN FOXP2.  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
OX NCBI\_TaxID=9598;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22179809; PubMed=12192408;  
RA Enard W., Przeworski M., Fisher S.E., Lai C.S.L., Wiebe V., Kitano T.,  
RA Monaco A.P., Paabo S.;  
RT "Molecular evolution of FOXP2, a gene involved in speech and  
RT language.";  
RL Nature 418:869-872(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22412141; PubMed=12524352;  
RA Zhang J., Webb D.M., Podlaha O.;  
RT "Accelerated protein evolution and origins of human-specific features:  
RT Foxp2 as an example.";  
RL Genetics 162:1825-1835(2002).  
CC -!- FUNCTION: Transcriptional repressor that plays an important role  
CC in the specification and differentiation of lung epithelium. May  
CC play important roles in developing neural, gastrointestinal and  
CC cardiovascular tissues (By similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- SIMILARITY: Contains 1 fork-head domain.  
CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF512947; AAN03385.1; -.  
DR EMBL; AF515051; AAN03409.1; -.  
DR EMBL; AF515052; AAN03410.1; -.  
DR EMBL; AY143178; AAN60056.1; -.  
DR InterPro; IPR001766; TF\_Fork\_head.  
DR InterPro; IPR009058; Wing\_hlx\_DNA\_bnd.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00250; Fork\_head; 1.  
DR PRINTS; PR00053; FORKHEAD.  
DR ProDom; PD000425; TF\_Fork\_head; 1.  
DR SMART; SM00339; FH; 1.  
DR PROSITE; PS00657; FORK\_HEAD\_1; FALSE\_NEG.

DR PROSITE; PS00658; FORK\_HEAD\_2; FALSE\_NEG.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; FALSE\_NEG.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein.  
 FT ZN\_FING 347 372 C2H2-TYPE.  
 FT DNA\_BIND 505 595 FORK-HEAD.  
 FT DOMAIN 53 56 POLY-GLN.  
 FT DOMAIN 123 126 POLY-GLN.  
 FT DOMAIN 131 136 POLY-GLN.  
 FT DOMAIN 152 191 POLY-GLN.  
 FT DOMAIN 201 210 POLY-GLN.  
 FT DOMAIN 224 232 POLY-GLN.  
 SQ SEQUENCE 716 AA; 80061 MW; 3169A2786B42F79F CRC64;

Query Match 49.3%; Score 187; DB 1; Length 716;  
 Best Local Similarity 97.4%; Pred. No. 4.5e-09;  
 Matches 37; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 15 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHG 52  
 |||  
 Db 158 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHPG 195

# RESULT 10

## LUG\_ARATH

ID LUG\_ARATH STANDARD; PRT; 931 AA.  
 AC Q9FUY2; Q9SZY9;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Transcriptional co-repressor LEUNIG.  
 GN LUG OR AT4G32551 OR L23H3.30.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=cv. Landsberg erecta;  
 RX MEDLINE=20524099; PubMed=11058164;  
 RA Conner J., Liu Z.;  
 RT "LEUNIG, a putative transcriptional corepressor that regulates AGAMOUS  
 RT expression during flower development.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12902-12907(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
 RA Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,

RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
 RA Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,  
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Granderrath K., Dauner D., Herzl A.,  
 RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,  
 RA Massenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
 RA Chedford F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,  
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M.A., Martienssen R., McCombie W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant *Arabidopsis*  
 RT *thaliana*.";  
 RL Nature 402:769-777(1999).  
 RN [3]  
 RP FUNCTION.  
 RX MEDLINE=95262573; PubMed=7743940;  
 RA Liu Z., Meyerowitz E.M.;  
 RT "LEUNIG regulates AGAMOUS expression in *Arabidopsis* flowers.";  
 RL Development 121:975-991(1995).  
 RN [4]  
 RP FUNCTION.  
 RX MEDLINE=21642140; PubMed=11782418;  
 RA Franks R.G., Wang C., Levin J.Z., Liu Z.;  
 RT "SEUSS, a member of a novel family of plant regulatory proteins,  
 RT represses floral homeotic gene expression with LEUNIG.";  
 RL Development 129:253-263(2002).  
 CC -!- FUNCTION: Acts as transcriptional co-repressor of the C class  
 CC floral homeotic gene AGAMOUS during the early stages of floral  
 CC meristem development. Is part of the A class cadastral complex  
 CC that define the boundaries between the A and C class homeotic



CC genes expression and function. Interacts together with APETALA2  
 CC and SEUSS to repress AGAMOUS expression. Also plays a role in  
 CC ovule and pollen development.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: Expressed in young flower primordia, later  
 CC becomes localized to petals, stamens and carpels. Is also  
 CC expressed in vegetative tissues.  
 CC -!- DEVELOPMENTAL STAGE: Expressed prominently during both female and  
 CC male gametes development.  
 CC -!- MISCELLANEOUS: Mutations in the LEUNIG gene result in the ectopic  
 CC expression of AGAMOUS, leading to the replacement of sepals by  
 CC carpels and stamens and of petals by stamens.  
 CC -!- SIMILARITY: Contains 1 Lish domain.  
 CC -!- SIMILARITY: Contains 7 WD repeats.  
 CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous  
 CC gene model prediction.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF277458; AAG32022.1; -.  
 DR EMBL; AL050398; CAB43692.1; ALT\_SEQ.  
 DR EMBL; AL161581; CAB79972.1; ALT\_SEQ.  
 DR TRANSFAC; T04599; -.  
 DR InterPro; IPR006594; Lish.  
 DR InterPro; IPR007591; SSDP.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF04503; SSDP; 1.  
 DR Pfam; PF00400; WD40; 7.  
 DR PRINTS; PR00320; GPROTEINBRPT.  
 DR SMART; SM00667; Lish; 1.  
 DR SMART; SM00320; WD40; 6.  
 DR PROSITE; PS50896; LISH; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 5.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Flowering; Transcription regulation; Repressor; Developmental protein;  
 KW Nuclear protein; Repeat; WD repeat.  
 FT DOMAIN 8 40 LISH.  
 FT REPEAT 641 679 WD 1.  
 FT REPEAT 683 721 WD 2.  
 FT REPEAT 726 765 WD 3.  
 FT REPEAT 769 804 WD 4.  
 FT REPEAT 808 845 WD 5.  
 FT REPEAT 852 890 WD 6.  
 FT REPEAT 893 931 WD 7.  
 FT DOMAIN 89 184 GLN-RICH.  
 FT DOMAIN 350 385 GLN-RICH.  
 FT DOMAIN 449 492 GLN-RICH.  
 FT CONFLICT 404 404 T -> S (IN REF. 1).  
 FT CONFLICT 454 454 N -> H (IN REF. 1).  
 SQ SEQUENCE 931 AA; 102232 MW; 7CB879744496A8AA CRC64;

Query Match 49.3%; Score 187; DB 1; Length 931;  
 Best Local Similarity 64.4%; Pred. No. 5.6e-09;  
 Matches 38; Conservative 2; Mismatches 17; Indels 2; Gaps 1;

```
Qy      11 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQQ--QQQQQQQQQQHHGNSGPPEFPGRLERPH 67
          |||||
Db      132 HHHHQQQQQQQQQQQQQQQQQQQQQQQQQHONQPPSQQQQQQSTPQHQQQPTPQQQPQRRDGS 190
```

RESULT 11

NIT4\_NEUCR

```
ID  NIT4_NEUCR      STANDARD;      PRT;  1090 AA.
AC  P28349;
DT  01-DEC-1992 (Rel. 24, Created)
DT  01-DEC-1992 (Rel. 24, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  Nitrogen assimilation transcription factor nit-4.
GN  NIT-4.
OS  Neurospora crassa.
OC  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC  Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX  NCBI_TaxID=5141;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92017855; PubMed=1840634;
RA  Yuan G.-F., Fu Y.-H., Marzluf G.A.;
RT  "nit-4, a pathway-specific regulatory gene of Neurospora crassa,
RT  encodes a protein with a putative binuclear zinc DNA-binding domain.";
RL  Mol. Cell. Biol. 11:5735-5745(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92149315; PubMed=1531376;
RA  Yuan G.-F., Marzluf G.A.;
RT  "Molecular characterization of mutations of nit-4, the pathway-
RT  specific regulatory gene which controls nitrate assimilation in
RT  Neurospora crassa.";
RL  Mol. Microbiol. 6:67-73(1992).
CC  -!- FUNCTION: PATHWAY-SPECIFIC REGULATORY GENE OF NITRATE
CC      ASSIMILATION; IT ACTIVATES THE TRANSCRIPTION OF THE GENES FOR
CC      NITRATE AND NITRITE REDUCTASES.
CC  -!- SUBCELLULAR LOCATION: Nuclear.
CC  -!- DOMAIN: The glutamine-rich domain might function in activating
CC      gene expression.
CC  -!- SIMILARITY: Contains 1 Zn(2)-Cys(6) fungal-type binuclear cluster
CC      domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M80368; AAA33602.1; -.
DR  PIR; A41696; A41696.
```

DR HSSP; P07272; 1PYI.  
 DR TRANSFAC; T02845; -.  
 DR InterPro; IPR007219; Fungal\_trans.  
 DR InterPro; IPR001138; Fungi\_TrN.  
 DR Pfam; PF04082; Fungal\_trans; 1.  
 DR Pfam; PF00172; Zn\_clus; 1.  
 DR SMART; SM00066; GAL4; 1.  
 DR PROSITE; PS00463; ZN2\_CY6\_FUNGAL\_1; 1.  
 DR PROSITE; PS50048; ZN2\_CY6\_FUNGAL\_2; 1.  
 KW Transcription regulation; Activator; DNA-binding; Nuclear protein;  
 KW Zinc; Metal-binding; Nitrate assimilation.  
 FT DNA\_BIND 53 81 ZN(2)-CYS(6), FUNGAL-TYPE.  
 FT DOMAIN 121 139 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 213 229 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 429 450 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 672 754 PRO-RICH.  
 FT DOMAIN 755 859 GLN-RICH.  
 FT DOMAIN 992 1024 POLY-GLN.  
 FT CONFLICT 98 98 K -> KP (IN REF. 1).  
 FT CONFLICT 467 467 L -> S (IN REF. 1).  
 SQ SEQUENCE 1090 AA; 120244 MW; 881D89172EDD6114 CRC64;

Query Match 48.0%; Score 182; DB 1; Length 1090;  
 Best Local Similarity 61.8%; Pred. No. 1.7e-08;  
 Matches 42; Conservative 4; Mismatches 2; Indels 20; Gaps 3;

Qy 1 LVPRGSV-----STHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ-- 49  
 | |||:: || |:||||:|||||||  
 Db 971 LAPRGNIGGGGGGGGGST----GQRQQQQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQEA 1026  
  
 Qy 50 -----HHG 52  
 |||  
 Db 1027 NMFAYHHG 1034

# RESULT 12

## FXP1\_MOUSE

ID FXP1\_MOUSE STANDARD; PRT; 705 AA.  
 AC P58462;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Forkhead box protein P1 (Forkhead-related transcription factor 1).  
 GN FOXP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).  
 RC STRAIN=C57BL/6; TISSUE=Lung;  
 RX MEDLINE=21347947; PubMed=11358962;  
 RA Shu W., Yang H., Zhang L., Lu M.M., Morrissey E.E.;  
 RT "Characterization of a new subfamily of winged-helix/forkhead (Fox)  
 RT genes that are expressed in the lung and act as transcriptional  
 RT repressors.";  
 RL J. Biol. Chem. 276:27488-27497(2001).

```

CC  -!- FUNCTION: Transcriptional repressor that play an important role in
CC      the specification and differentiation of lung epithelium.
CC  -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Name=A;
CC          IsoId=P58462-1; Sequence=Displayed;
CC          Note=Isoform C is produced by alternative initiation at Met-251
CC            of isoform A;
CC      Name=B;
CC          IsoId=P58462-2; Sequence=VSP_001557;
CC      Event=Alternative initiation;
CC          Comment=2 isoforms, A (shown here) and C, are produced by
CC            alternative initiation at Met-1 and Met-251;
CC  -!- TISSUE SPECIFICITY: Highest expression in the lung, brain, and
CC      spleen. Lower expression in heart, skeletal muscle, kidney, small
CC      intestine (isoform C not present) and liver.
CC  -!- DEVELOPMENTAL STAGE: Expressed in developing lung, neural,
CC      intestinal and cardiovascular tissues.
CC  -!- SIMILARITY: Contains 1 fork-head domain.
CC  -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AF339103; AAK69648.1; -.
DR  EMBL; AF339104; AAK69649.1; -.
DR  EMBL; AF339105; AAK69650.1; -.
DR  MGD; MGI:1914004; Foxp1.
DR  GO; GO:0016564; F:transcriptional repressor activity; IDA.
DR  GO; GO:0016481; P:negative regulation of transcription; IDA.
DR  InterPro; IPR001766; TF_Fork_head.
DR  InterPro; IPR007087; ZnF_C2H2.
DR  Pfam; PF00250; Fork_head; 1.
DR  PRINTS; PR00053; FORKHEAD.
DR  ProDom; PD000425; TF_Fork_head; 1.
DR  SMART; SM00339; FH; 1.
DR  SMART; SM00355; ZnF_C2H2; 1.
DR  PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR  PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR  PROSITE; PS50039; FORK_HEAD_3; 1.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR  PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW  Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW  Nuclear protein; Alternative splicing; Alternative initiation.
FT  CHAIN      1      705      FORKHEAD BOX PROTEIN P1, ISOFORM A.
FT  CHAIN      251    705      FORKHEAD BOX PROTEIN P1, ISOFORM C.
FT  INIT_MET    251    251      FOR ISOFORM C.
FT  DNA_BIND    493    583      FORK-HEAD.
FT  ZN_FING     334    359      C2H2-TYPE.
FT  DOMAIN      55     60      POLY-GLN.
FT  DOMAIN      71    107      POLY-GLN.

```

FT DOMAIN 161 164 POLY-GLN.  
 FT VARSPLIC 539 602 Missing (in isoform B).  
 FT /FTid=VSP\_001557.  
 SQ SEQUENCE 705 AA; 78833 MW; 92962B82917CC79D CRC64;

Query Match 47.8%; Score 181; DB 1; Length 705;  
 Best Local Similarity 88.1%; Pred. No. 1.4e-08;  
 Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 15 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGP 56  
 |||||||||||||||||||||||||||||||| | |  
 Db 73 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQVSGLKSP 114

# RESULT 13

## YM38\_YEAST

ID YM38\_YEAST STANDARD; PRT; 758 AA.  
 AC Q03825;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical 85.0 kDa protein in HLJ1-SMP2 intergenic region.  
 GN YMR164C OR YM8520.13C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX MEDLINE=97313268; PubMed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT XIII.";  
 RL Nature 387:90-93(1997).  
 CC -!- SIMILARITY: Contains 1 LisH domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z49705; CAA89800.1; -.  
 DR PIR; S54522; S54522.  
 DR GermOnline; 142836; -.  
 DR SGD; S0004774; MSS11.  
 DR GO; GO:0005634; C:nucleus; IC.  
 DR GO; GO:0003704; F:specific RNA polymerase II transcription fa. . .; IDA.  
 DR GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.  
 DR GO; GO:0007124; P:pseudohyphal growth; IGI.  
 DR GO; GO:0005983; P:starch catabolism; IMP.



RP SEQUENCE OF 189-2212 FROM N.A.  
 RX MEDLINE=98368120; PubMed=9702738;  
 RA Philibert R.A., King B.H., Cook E.H., Lee Y.-H., Stubblefield B.,  
 RA Damschroder-Williams P., Dea C., Palotie A., Tengstrom C.,  
 RA Martin B.M., Ginns E.I.;  
 RT "Association of an X-chromosome dodecamer insertional variant allele  
 RT with mental retardation.";  
 RL Mol. Psych. 3:303-309(1998).  
 RN [4]  
 RP SEQUENCE OF 189-2212 FROM N.A.  
 RX MEDLINE=99408253; PubMed=10480376;  
 RA Philibert R.A., Winfield S.L., Damschroder-Williams P., Tengstrom C.,  
 RA Martin B.M., Ginns E.I.;  
 RT "The genomic structure and developmental expression patterns of the  
 RT human OPA-containing gene (HOPA).";  
 RL Hum. Genet. 105:174-178(1999).  
 RN [5]  
 RP SEQUENCE OF 1564-2212 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97369492; PubMed=9225980;  
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,  
 RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;  
 RT "cDNAs with long CAG trinucleotide repeats from human brain.";  
 RL Hum. Genet. 100:114-122(1997).  
 RN [6]  
 RP IDENTIFICATION IN ARC COMPLEX, AND SEQUENCE OF 1709-1717 AND  
 RP 1806-1817.  
 RX MEDLINE=99249346; PubMed=10235267;  
 RA Naeaeer A.M., Beaurang P.A., Zhou S., Abraham S., Solomon W.B.,  
 RA Tjian R.;  
 RT "Composite co-activator ARC mediates chromatin-directed  
 RT transcriptional activation.";  
 RL Nature 398:828-832(1999).  
 CC -!- FUNCTION: Plays a role in transcriptional coactivation.  
 CC -!- SUBUNIT: Subunit of the large multiprotein complexes TRAP and  
 CC ARC/DRIP.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: Ubiquitous.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF117755; AAD22033.1; -.  
 DR EMBL; D83783; BAA12112.1; -.  
 DR EMBL; AF071309; AAC83163.1; -.  
 DR EMBL; AF132033; AAD44162.1; -.  
 DR EMBL; U80742; AAB91440.1; -.  
 DR Genew; HGNC:11957; TNRC11.  
 DR MIM; 300188; -.  
 DR GO; GO:0000119; C:mediator complex; IDA.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR GO; GO:0030374; F:ligand-dependent nuclear receptor transcrip. . .; NAS.

DR GO; GO:0004872; F:receptor activity; IDA.  
 DR GO; GO:0016455; F:RNA polymerase II transcription mediator ac. . .; IDA.  
 DR GO; GO:0046966; F:thyroid hormone receptor binding; IDA.  
 DR GO; GO:0016563; F:transcriptional activator activity; IDA.  
 DR GO; GO:0042809; F:vitamin D receptor binding; NAS.  
 DR GO; GO:0030521; P:androgen receptor signaling pathway; IDA.  
 DR GO; GO:0006367; P:transcription initiation from Pol II promoter; IDA.  
 KW Transcription regulation; Activator; Receptor; Nuclear protein.  
 FT DOMAIN 1289 1295 POLY-GLY.  
 FT DOMAIN 2086 2212 GLN-RICH.  
 FT DOMAIN 2086 2111 POLY-GLN.  
 FT DOMAIN 2116 2121 POLY-GLN.  
 FT DOMAIN 2125 2158 POLY-GLN.  
 FT DOMAIN 2178 2185 POLY-GLN.  
 FT CONFLICT 1201 1201 E -> V (IN REF. 4).  
 FT CONFLICT 1427 1427 R -> Q (IN REF. 3 AND 4).  
 FT CONFLICT 1951 1951 MISSING (IN REF. 3 AND 4).  
 FT CONFLICT 1951 1951 Q -> QAKI (IN REF. 5).  
 SQ SEQUENCE 2212 AA; 247333 MW; E959525836147630 CRC64;

Query Match 47.6%; Score 180.5; DB 1; Length 2212;  
 Best Local Similarity 71.2%; Pred. No. 4.2e-08;  
 Matches 37; Conservative 3; Mismatches 9; Indels 3; Gaps 1;

Qy 10 HHHHHQQQQQ---QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHHGNSGPPE 58  
 :| |||| | ||||| ||||| ||||| ||||| ||||| ||||| : ||:  
 Db 2112 YHIRQQQQQILRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHQQQQQQQAAPPQ 2163

# RESULT 15

## TBP\_HUMAN

ID TBP\_HUMAN STANDARD; PRT; 339 AA.  
 AC P20226; Q16845; Q9UC02;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE TATA-box binding protein (TATA-box factor) (TATA binding factor) (TATA  
 DE sequence-binding protein) (Transcription initiation factor TFIID TBP  
 DE subunit).  
 GN TBP OR TFIID OR TF2D.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND DOMAINS.  
 RX MEDLINE=90302006; PubMed=2363050;  
 RA Peterson M.G., Tanese N., Pugh B.F., Tjian R.;  
 RT "Functional domains and upstream activation properties of cloned  
 RT human TATA binding protein."  
 RL Science 248:1625-1630(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=90302010; PubMed=2194289;  
 RA Kao C.C., Lieberman P.M., Schmidt M.C., Zhou Q., Pei R., Berk A.J.;  
 RT "Cloning of a transcriptionally active human TATA binding factor.";



RL Science 248:1646-1650(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND VARIANT 92-GLN--GLN-95 DEL.  
 RX MEDLINE=90326195; PubMed=2374612;  
 RA Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M.,  
 RA Roeder R.G.;  
 RT "Highly conserved core domain and unique N terminus with presumptive  
 RT regulatory motifs in a human TATA factor (TFIID).";  
 RL Nature 346:387-390(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Griffiths C.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP INTERACTION WITH NCOA6.  
 RX MEDLINE=20036574; PubMed=10567404;  
 RA Lee S.-K., Anzick S.L., Choi J.-E., Bubendorf L., Guan X.-Y.,  
 RA Jung Y.-K., Kallioniemi O.P., Kononen J., Trent J.M., Azorsa D.,  
 RA Jhun B.-H., Cheong J.H., Lee Y.C., Meltzer P.S., Lee J.W.;  
 RT "A nuclear factor ASC-2, as a cancer-amplified transcriptional  
 RT coactivator essential for ligand-dependent transactivation by nuclear  
 RT receptors in vivo.";  
 RL J. Biol. Chem. 274:34283-34293(1999).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337 IN COMPLEX WITH DNA.  
 RX MEDLINE=96209823; PubMed=8643494;  
 RA Nikolov D.B., Chen H., Halay E.D., Hoffmann A., Roeder R.G.,  
 RA Burley S.K.;  
 RT "Crystal structure of a human TATA box-binding protein/TATA element  
 RT complex.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4862-4867(1996).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339 IN COMPLEX WITH DNA.  
 RX MEDLINE=96346176; PubMed=8757291;  
 RA Juo Z.S., Chiu T.K., Leiberman P.M., Baikarov I., Berk A.J.,  
 RA Dickerson R.E.;  
 RT "How proteins recognize the TATA box.";  
 RL J. Mol. Biol. 261:239-254(1996).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.65 ANGSTROMS) OF 159-337 IN COMPLEX WITH  
 RP GTF2B AND DNA.  
 RX MEDLINE=20086817; PubMed=10619841;  
 RA Tsai F.T.F., Sigler P.B.;  
 RT "Structural basis of preinitiation complex assembly on human pol II  
 RT promoters.";  
 RL EMBO J. 19:25-36(2000).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.62 ANGSTROMS) OF 159-339 IN COMPLEX WITH DR1;  
 RP DRAP1 AND DNA.  
 RX MEDLINE=21354312; PubMed=11461703;  
 RA Kamada K., Shu F., Chen H., Malik S., Stelzer G., Roeder R.G.,  
 RA Meisterernst M., Burley S.K.;  
 RT "Crystal structure of negative cofactor 2 recognizing the TBP-DNA  
 RT transcription complex.";  
 RL Cell 106:71-81(2001).  
 RN [10]  
 RP POLYMORPHISM OF POLY-GLN REGION.

RX MEDLINE=99415745; PubMed=10484774;  
 RA Koide R., Kobayashi S., Shimohata T., Ikeuchi T., Maruyama M.,  
 RA Saito M., Yamada M., Takahashi H., Tsuji S.;  
 RT "A neurological disease caused by an expanded CAG trinucleotide repeat  
 RT in the TATA-binding protein gene: a new polyglutamine disease?";  
 RL Hum. Mol. Genet. 8:2047-2053(1999).  
 RN [11]  
 RP POLYMORPHISM OF POLY-GLN REGION.  
 RX MEDLINE=21214723; PubMed=11313753;  
 RA Zuhlke C., Hellenbroich Y., Dalski A., Kononowa N., Hagenah J.,  
 RA Vieregge P., Riess O., Klein C., Schwinger E.;  
 RT "Different types of repeat expansion in the TATA-binding protein gene  
 RT are associated with a new form of inherited ataxia.";  
 RL Eur. J. Hum. Genet. 9:160-164(2001).  
 RN [12]  
 RP POLYMORPHISM OF POLY-GLN REGION.  
 RX MEDLINE=21341926; PubMed=11448935;  
 RA Nakamura K., Jeong S.-Y., Uchihara T., Anno M., Nagashima K.,  
 RA Nagashima T., Ikeda S.-I., Tsuji S., Kanazawa I.;  
 RT "SCA17, a novel autosomal dominant cerebellar ataxia caused by an  
 RT expanded polyglutamine in TATA-binding protein.";  
 RL Hum. Mol. Genet. 10:1441-1448(2001).  
 RN [13]  
 RP POLYMORPHISM OF POLY-GLN REGION.  
 RX MEDLINE=21937712; PubMed=11939898;  
 RA Silveira I., Miranda C., Guimaraes L., Moreira M.-C., Alonso I.,  
 RA Mendonca P., Ferro A., Pinto-Basto J., Coelho J., Ferreira F.,  
 RA Poirier J., Parreira E., Vale J., Januario C., Barbot C., Tuna A.,  
 RA Barros J., Koide R., Tsuji S., Holmes S.E., Margolis R.L., Jardim L.,  
 RA Pandolfo M., Coutinho P., Sequeiros J.;  
 RT "Trinucleotide repeats in 202 families with ataxia: a small expanded  
 RT (CAG)<sub>n</sub> allele at the SCA17 locus.";  
 RL Arch. Neurol. 59:623-629(2002).  
 CC -!- FUNCTION: General transcription factor that functions at the  
 CC core of the DNA-binding multiprotein factor TFIID. Binding of  
 CC TFIID to the TATA box is the initial transcriptional step of the  
 CC pre-initiation complex (PIC), playing a role in the activation of  
 CC eukaryotic genes transcribed by RNA polymerase II.  
 CC -!- SUBUNIT: Belongs to the TFIID complex together with the TBP-  
 CC associated factors (TAFs). Binds DNA as monomer. Interacts with  
 CC TAFs, TFIIA, TFIIB, NCOA6, DRAP1 and DR1.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- POLYMORPHISM: The poly-Gln region of TBP is highly polymorphic (25  
 CC to 42 repeats) in normal individuals and is expanded to about 47-  
 CC 63 repeats in SCA17 patients. Longer expansions may result in  
 CC earlier onset and more severe clinical manifestations of the  
 CC disease.  
 CC -!- DISEASE: Defects in TBP are the cause of spinocerebellar ataxia  
 CC type 17 (SCA17) [MIM:607136]. SCA17 is a rare autosomal dominant  
 CC neurodegenerative disease, characterized by gait ataxia and  
 CC dementia, progressing over several decades to include  
 CC bradykinesia, dysmetria, dysidiadochokinesis, hyperreflexia and  
 CC paucity of movement.  
 CC -!- SIMILARITY: Belongs to the TBP family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M55654; AAA36731.1; -.  
 DR EMBL; M34960; AAC03409.1; -.  
 DR EMBL; X54993; CAA38736.1; -.  
 DR EMBL; AL031259; CAA20286.1; -.  
 DR PIR; A34830; TWHU2D.  
 DR PDB; 1CDW; 23-DEC-96.  
 DR PDB; 1C9B; 10-JAN-00.  
 DR PDB; 1JFI; 11-JUL-01.  
 DR PDB; 1TGH; 01-AUG-96.  
 DR TRANSFAC; T00794; -.  
 DR Genew; HGNC:11588; TBP.  
 DR MIM; 600075; -.  
 DR MIM; 607136; -.  
 DR GO; GO:0005669; C:transcription factor TFIID complex; TAS.  
 DR GO; GO:0016251; F:general RNA polymerase II transcription fac. . .; TAS.  
 DR GO; GO:0006367; P:transcription initiation from Pol II promoter; TAS.  
 DR InterPro; IPR000814; TFIID.  
 DR Pfam; PF00352; TBP; 2.  
 DR PRINTS; PR00686; TIFACTORIID.  
 DR PROSITE; PS00351; TFIID; 2.  
 KW Transcription; Nuclear protein; DNA-binding; Repeat; Polymorphism;  
 KW Triplet repeat expansion; Disease mutation; 3D-structure.  
 FT REPEAT 165 241 1.  
 FT REPEAT 255 332 2.  
 FT DOMAIN 55 95 POLY-GLN.  
 FT VARIANT 92 95 Missing.  
 FT /FTId=VAR\_016987.  
 FT CONFLICT 187 187 A -> R (IN REF. 2).  
 FT STRAND 164 173  
 FT HELIX 180 186  
 FT TURN 188 189  
 FT STRAND 190 192  
 FT TURN 194 196  
 FT STRAND 199 205  
 FT TURN 206 207  
 FT STRAND 208 214  
 FT TURN 215 216  
 FT STRAND 218 222  
 FT HELIX 227 244  
 FT TURN 245 245  
 FT STRAND 251 263  
 FT STRAND 268 268  
 FT HELIX 270 276  
 FT TURN 277 280  
 FT STRAND 281 282  
 FT TURN 285 287  
 FT STRAND 291 295  
 FT TURN 296 299  
 FT STRAND 300 304  
 FT TURN 306 307  
 FT STRAND 309 313

FT HELIX 318 333  
FT TURN 334 335  
FT STRAND 336 336

Query Match 47.2%; Score 179; DB 1; Length 339;  
Best Local Similarity 81.8%; Pred. No. 1.1e-08;  
Matches 36; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 6 SVSTHHHHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 49  
|: |  
Db 50 SILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 93

Search completed: March 12, 2004, 15:39:05  
Job time : 9.11765 secs